To Look, But Not to Learn:

The Impact of Televised Formal Features on Attention and Comprehension of Infant Learning Paradigms

Marcos Sastre III

A Senior Honors Thesis for the Department of Psychology

Tufts University

Abstract

Young infants (under 2 years old) are widely underrepresented in media research, a fact that has spurned numerous debates on the effects of screen viewing on development. Past literature has indicated that the degree of perceptual salience on video is the most robust mediator of infant attention. Nevertheless, attention is not a reliable predictor of comprehension by itself; rather, the different audiovisual properties of the stimulus can have varying effects on processing information from a video source. The current study aims to isolate how formal features, 'syntactic' elements of design and production, affect attention and comprehension of video demonstrations on two forced choice imitation/search tasks. Infants were shown a series of six short video clips, each modified with one of four experimental conditions (two visual, two auditory): Cartoon Voice, Music, Visual Effects, Artificial Dynamics, and Music, and two control trials.

Results indicated that unexpectedly, attention to the medium was high overall, with few condition effects (Hypothesis 1), but, as predicted, imposition of formal features resulted in poorer performance on all trials except the Cartoon Voice condition (Hypothesis 2). Condition differences were also observed for task performance, as expected (Hypothesis 3). These findings suggest that children are capable of learning from video, but this ability is markedly affected by the audiovisual conventions that often accompany commercial products. The implications of these findings in light of the controversies surrounding infant-direct media are discussed.

Keywords: infants, media, attention, television, learning from video, formal features

The Impact of Televised Formal Features on Attention and Comprehension of Infant Learning Paradigms

The last decade of commercial and technological progress has drastically changed the media landscape of the infant viewer. Throughout the 1990s and into the new millennium, the target consumer for media has trickled down to the youngest age groups. Forty years ago, the most seminal work in children's media was being completed on elementary age perceptions of brand and selling intent; today, the market for baby "learning" media has grown astronomically, with numerous commercially branded videos and DVDs claiming the ability to optimize infant development (Strasburger, Wilson, & Jordan, 2009). Thus, as target consumers have gotten younger, more awareness has been placed on the amount and type of media interaction that some infants as young as 3 months old find themselves engaging in every day.

The American Academy of Pediatrics (1999) recommends that children under 2 years old should not be exposed to television or screen media of any kind, especially in a commercial context. Nonetheless, the most recent comprehensive studies on young child viewing behaviors found that by 3 months old, 40% of American children had already begun to watch television on a daily basis, and by 24 months, this number had risen to 90%, with a median age of first exposure around 9 months (Zimmerman, Christakis, & Meltzoff, 2007). Zimmerman et al. (2007) also reported how parents estimated their own children between 2 and 24 months as indulging in an average daily viewing time of 40.2 minutes per day, while Rideout and Hamel (2006) reported that by age 2, this number rises to 120 minutes per day. An observational study by Courage and Setliff (2010) determined that 19% of toddlers (0 through 3) have a television in their bedroom; 38% are capable of turning it on themselves, and 40% can change the channel.

The range of these media sources is widespread; Zimmerman et al. (2007) reported how that infants who have two or more siblings are more likely to be exposed to more advertisements to those with no siblings. But perhaps the fastest growing segment of the infant media market in recent years has been infant learning videos, which promise the abilities to enhance cognitive development and hone a variety of different skills. A $200 million dollar industry, the infant video market, pioneered by well-known brands such as *Baby Einstein* and *Brainy Baby*, has faced many controversies from advocacy groups attempting to accuse these marketers of false advertising and causing harm to infant development (Courage, Murphy, Goulding, & Setliff, 2010). Furthermore, the growth of 24-hour infant learning channels, such as *BabyFirst TV* and *Sprout*, has fueled these rebuttals even further. A small corpus of studies has tried to link longitudinal benefits (from better readiness for school to language learning and development of social skills) or detriment (such as attention difficulties, reading deficits, and replacing other activities) with few convincing arguments either way (Courage and Setliff, 2010). Nonetheless, the general academic consensus has been that at this point, not enough research exists to establish a blanket valence over what effect these videos have (Barr, Zack, Garcia, & Muentener, 2008; Robb, Richert, & Wartella, 2009).

In the last forty years, research observing the diverse variables present to the infant viewer has grown. Thus, the larger question of ‘learning from video’ is addressed, and the degree to which it compares and contrasts to learning from live models is a highly intriguing phenomenon.

It has become well established that, as infants grow, the real-world knowledge they acquire influences their comprehension of current content (Anderson & Hanson, 2010; Huston & Wright, 1983; Wright et al., 1984). In contrast, television, despite its content per se, is ridden with conventions that set it apart from real-world experience of the same stimuli. The use of animation, voice-overs, any kinds of visual or sound effects, pacing changes, zooms, and camera angles, are all examples of phenomena known as televised formal features. Huston and Wright (1983) loosely define formal features as the use of various audiovisual devices unique to a particular medium that, as attributes separate from content, are relevant to the program. Work with these various features has garnered a range of opinions on their purpose and use; Anderson and Bryant (1983), and Huston and Wright (1983) proposed that overall, formal features have functions and elicit effects independent from content, and thus are not arbitrary embellishments. Bickham, Wright, and Huston (2001) take this idea further, stressing that the use of formal features are a function of intended age; namely, effective use of formal features depends on whether or not they serve the developmental needs of the viewer and work with content to enhance it, not obfuscate it.

Relevance of the Target Age Group

Overall, the infant (0 to 2 years) and toddler (2 to 4 years) age groups have been given the least amount of study with media sources relative to older children. Until recently, infants were not a targeted age group for producers of commercial branded content (namely, videos, DVDs, and programs). Furthermore disagreement on how to categorize programming content by age has made judging ‘age-appropriateness’ difficult

(Christakis & Zimmerman, 2009). Therefore, despite the materials themselves, younger viewers have more developmental limitations, and will most likely misinterpret the message being communicated (Beentjes, de Koning, and Huysmans, 2001). These limitations correspond primarily to the lack of experience with both media and the real world (Christakis & Zimmerman, 2009). How are these limitations defined, and how do formal features influence these variables?

In comparison to other age groups, the most relevant defining factor about infant viewership is that from 0 to 2 years, observance of media is defined primarily by an overreliance on perceptual salience (Strasburger et al., 2009). Anderson and Hanson (2010) cite how even the youngest neonates (0 to 6 months old), who have poorly developed visual acuity, still orient to brighter on screen by six months, and salient auditory responses occur even earlier. Ideas on perceptual fixedness stems from Piaget's (1930) applied developmental stage theory; 0 to 2 year-olds fall into the earliest sensorimotor stage, where they are still entirely focused on the perceptual properties of stimuli, and their comprehension is limited by centration, or the ability to only focus on one feature at a time. Valkenburg and Cantor (2001) expand on this idea to include how attention to salience is a measure of discovering preferences, fostering attraction to bright colors, sweet tastes, and rhythmic auditory stimuli.

Complexity in a stimulus is associated with its level of salience. Wartella & Ettema (1974) have found that highly complex visual stimuli, categorized by visual diversity and entropy (randomized motion), are preferred by nursery age children. Sudden changes in the perceptual field, such as a loud noise or a change of pace have been found to elicit an orienting response to the screen (Gunter and McAleer, 1990), exemplifying the infants' preference for novel items (Anderson, Levin, & Lorch, 1977; Barr et al., 2008; Murray & Murray, 2008). Overall, however, the preference for salience has implications beyond attention, in that it may affect the interpretation of content. Hoffner and Cantor (1985) conducted a study observing preference (via attention) for an animated character based on a matrix of design vs. action. Four pictures were shown: a rounded, smiling woman treating a cat nicely, the same woman treating the cat harshly, and a wizened, frightening-looking woman completing both actions separately. The researchers found that 3-year-olds were much more likely to prefer the round and pleasant-looking woman to the frightening-looking one regardless of action compared to 6-year-olds, who preferred the good behavior regardless of appearance; such a finding affirms that a fixation on the surface features of stimuli obfuscates a deeper understanding of other variables in younger children (Hoffner & Cantor, 1985).

Thus, this notion of fixation and preference for salience creates many misconceptions about the nature of comprehension. One school of thought promotes the passive model of viewership, which suggests that infants ‘mindlessly’ absorb stimuli (Bickham et al., 2001, Strasburger et al., 2009). In contrast, the active model of viewership, pioneered by Anderson, Lorch, Field, and Sanders (1981), suggests that children are actually physically and cognitively engaged in what occurs on the screen. Lemish (2007) cites a summary of qualitative accounts of infants 0 to 2 exploring during their television viewing, mentioning how babies would laugh and clap to music, attempt to touch shapes and forms depicted, and “explore differences” (p. 38) between what they witnessed on screen and what they were used to seeing in the real world. This can be thought of as the forming of active schemas that help establish familiarity with the audiovisual conventions of the content itself (Lemish, 2007). The active model portrays media consumption in a positive light, such that viewing sharpens to tools to understand cues from media more fully as the child grows.

Other theories of salience fixation fall between the active/passive dichotomy, stressing the limitations of comprehension despite active engagement; Roedder's (1981) theory of information processing cites how children under the age of 7 are incapable of successfully diverging larger context from what they are viewing, despite their active investment. Another such theory is the exploration search theory (Huston & Wright, 1983), which emphasizes how the fixation on perceptual salience motivates exploration of the stimulus more thoroughly; in contrast, older children engage in searching, which causes them to selectively orient towards content they already know they will find more interesting (Huston & Wright, 1983; Schmitt, Anderson, & Collins, 1999). Lastly, Valkenburg and Vroone (2004) have proposed the moderate discrepancy hypothesis, which suggests that children attend to salience because doing so fits accurately within their developmental ability. This theory postulates that, as infants acquire the cognitive tools necessary to interpret more complex content, their exclusive preference for salience diminishes accordingly (Valkenburg & Vroone, 2004).

Attention to Television

Thus, the infant fixation on perceptual salience can be used to inform studies on attention to on-screen stimuli (Strasburger et al., 2009). As Valkenburg and Vroone (2004) mention, this orienting response delineates what the infant finds engaging and interesting because it is novel— once the child habituates to the stimulus, he or she will lose interest. Thus, the idea of formal features dictating attention has become an intriguing question for researchers to answer, using both measures of looking time, as well as derived behavior from the actions themselves. Barr, Lauricella, and Zack (2010a), found that a sample of 12- to 18-month-olds averaged an overall looking time to unmodified, on-screen stimuli of approximately 70%. Seehagen and Herbert (2009), in their televised 'learning from voices' paradigm, found even higher attention across conditions of 81.83%, on average. While looking time has been almost ubiquitously used as a measure of attention, Richards (2010) designed a demonstration that utilized change in heart rate as a measure of changes in visual stimuli, finding a prominent rise with the onset of a new stimulus, followed by a steady decline and plateau with habituation.

In observing age differences, this attention to on-screen stimuli follows a steady increase from 6 to 24 months (Richards, 2010). Gunter and McAleer's (1990) results also reflected this, finding a fourfold increase in looking time between 12 and 48 months, with the greatest increase at 30 months. In a cross-age sample observing attention to different modalities of televised stimuli (e.g. child-oriented programming, advertisements), only the 2-year-old group (the youngest) maintained attention to both modalities equally, and did so most readily when co-viewing with others (Schmitt, Woolf and Anderson, 2003); this same work cited how attention remained constant even when other toys and activities were present. Given this, Anderson, Choi, and Lorch (1979) introduced the idea of attentional inertia, which observes that the longer a child is able to attend to an on-screen stimulus, the less chance there is of that fixation being broken over time. Once the initial contact between the stimulus and the perceiver is formed, the fixation is less likely to be terminated than with older groups—this idea of a "hook" has various implications for discussion (Richards & Anderson, 2004; Richards, 2010). Nonetheless, many studies (Anderson & Levin, 1976; Wartella & Ettema, 1974) have confirmed that despite this, the longest uninterrupted fixation that a child younger than 2 years is capable of is no longer than 60 seconds, compared to four-year olds, who can attend to the screen as long as 7 minutes without looking up (Anderson & Levin, 1976).

The factors that mediate attention are vast and complex, but in recent years many theories have pointed to the role of comprehension and its correlation with attention. As expressed in the active model, children attend because they are seeking information derived from what they see in front of them. The stimulus sampling model of Huston and Wright (1983) champions this notion, expressing how as children gain viewership experience, they will use the formal features of programs to as markers of relevant content—in the infancy phase, the level of comprehension is thought to reflect the idea that 'if it is emphasized, it must be important,' regardless of whether or not the child understands why. As Gunter and McAleer (1990) succinctly expressed, visual attention is "largely governed by how much they understand what's going on" (p. 40).

Therefore, if the importance of salience is assumed, formal features designed for the infant viewer have a much more important role than arbitrary embellishment, rather they can be used as relevant markers of information for infants to attend to what's most important (e.g. a change, a new introduction, or emphasis on an action; Huston & Wright, 1983). This has practical (and potentially educational) implications, specifically in recognizing how formal features can be designed specifically for the infant viewer. Barr et al., (2010a) completed a study observing child viewership of children versus adult programming, finding that, both in attention and cognitive outcomes, infants were more ‘productive’ with content designed for their age group. Twelve to 18-month olds were estimated to attend for 70% of the time to children’s content and only 5% to non-infant content; furthermore, those categorized to be heavy viewers of non-infant content between 1 and 4 years were found to score less on assessments of general intelligence, vocabulary, and inhibition. Barr et al. (2010a) concluded that due to richer saturation of formal features in child-oriented content, children were able to glean more from their viewing experience over time, stressing that assessing content is just as important, if not more so, than assessing time spent in front of media to understand long-term effects.

Comprehension of Content

Therefore, comprehension of content, and the degree to which formal features influence this, is of vital importance to understanding infants’ interaction with video sources. Over time, numerous strategies and measures have developed for assessing the broad category of ‘learning from video, and the type of task used is relevant to discern which types of cognition are in effect. Imitation is an often-used paradigm in video studies. As a measure, Barr, Dowden and Hayne (1996) have assessed how imitation yields more effective outcomes than trial-and-error learning and spontaneous problem solving (sans stimulus—a method often used as the baseline control for imitation work), because it allows for assessment of the relationship between the stimulus and the task. Deferred imitation (24 hours after the original demonstration) includes these benefits, while assessing retention of content; it is assumed that if a child can imitate a task presented to him or her 24 hours before, they were able to encode and retain the information until the time of test (Barr et al., 1996). Meltzoff (1988) asserts that both forms of imitation are a relevant measure to assess pre-linguistic, representations of learning in a time where children cannot eloquently express what they understand.

Overall, the ability to imitate from video is age-dependent, with higher imitation ability scores increasing until 18 months (Barr et al., 1996) but plateauing until about 30 months for tasks from video content (Hayne, Herbert, & Simcock, 2003). Evidence of deferred imitation from video is found to not be present robustly before 12 months, though increasing the time of exposure to demonstrate a video paradigm can significantly improve the ability in 6-month-olds (Barr et al., 1996). A seminal study by McCall, Parke, and Kavanaugh (1977) demonstrated slightly more conservative figures, assessing that multi-step sequences of coordinated behavior (beyond ‘single unit,’ one action tasks) require at least 18 months of age for successful imitation and 24 for successful deferred imitation. They found that, for live models, imitation of single unit behaviors rose from 28% to 78% between 12 and 24 months; even for those recorded to be heavy television viewers, the researchers found no established relationship between reported home viewing and propensity to imitate more readily before 36 months (McCall et al., 1977). Overall, older infants have been judged to learn faster and remember longer than younger ones (Barr & Hayne, 1999). Furthermore, Hayne et al.’s (2003) work revealed that at 6 months, a change in the testing paradigm and testing context (location) between trials negatively influenced imitation; by 12 months, change in context was no longer disruptive, and by 18 months, neither was disruptive for imitation.

Another common measure of learning is the object search and recognition task, which asks infants not to copy what they learned from a source (live or televised), but to derive the nature of the task from the prompted information (i.e. find an object, the location of which is given in the demonstration). Much of Troseth's (2003a; 2003b; with DeLoache, 1998; with Saylor & Archer, 2006) work has utilized this paradigm specifically in placing the object in another room, and then having the child find that object based on the transfer of information from the stimulus to the real setting. Schmitt and Anderson (2002) determined that most 2-year-olds were able to find toys in a real room, based on a video-prompted search task that depicted the toy being hidden. Troseth and DeLoache (1996) express the relevance of using object search and recognition from video as a measure not only for transferring information between media, but also for making decisions in a new context based on the presented information.

Video Deficit Effect

Since comparisons between live and televised paradigms developed, it became clear that despite the ability to imitate and transfer from video (McCall et al., 1977; Meltzoff, 1988; others), this capacity is significantly diminished when the source of the prompt is televised relative to a live presentation. This concept has been termed the video deficit effect, and is discussed in a number of studies (Macklin, 1994; Murray & Murray, 2008; Barr, Wyss, & Somanader, 2009; Anderson & Hanson, 2010; Courage & Setliff, 2010). Barr and Hayne (1996) utilized an imitation task with 12- and 18-month-olds, where an action was performed on a target object, with the experimental group receiving a televised instructional prompt, and the control group receiving the same prompt with a live demonstrator. For both ages, infants in the live condition performed significantly better on imitating the target action than infants in the televised condition; there was an interaction effect with the older age group, which also was the only able to still complete the video imitation task when imitation was deferred for 24 hours. Even in these trials, live demonstrations yielded stronger imitation than direct imitation from video (Barr & Hayne, 1996).

The video deficit effect also was present for object-search tasks conducted by Troseth & DeLoache (1998), in which infants 24 and 36 months old watched information about a toy being hidden in the next room on a television; older infants had no problem finding the toy, achieving 79% errorless retrieval, and performed similarly when watching the same hiding event through a window (live demonstration condition); nonetheless, the younger group was not able to perform above baseline, finding the toy from the televised prompt on the first try (errorless retrieval) only 44% of the time in the direct imitation paradigm. Barr, Shuck, Salerno, Atkinson, and Linebarger (2010b) postulated that the video deficit effect peaks about 15 months for imitation tasks, and 24 months for object recognition, persisting only until about 3 years (Schmitt & Anderson (2002), and Anderson & Hanson (2010), agree). Some estimates, such as Flynn and Whiten (2008), are more conservative, encountering the video deficit as late as 5 years.

Interestingly, Troseth and DeLoache (1998) also found that knowledge of the medium was a prominently influential factor affecting performance; they noticed that when young children *believed* they were witnessing the activity through the window rather than the television (and the television was embedded and framed to look like a window and called the ‘video-window’ condition), they performed better on the video task, despite the perceptual similarity in the stimuli. Overall, 63% of 2.5 year olds who were actually watching a video, but were told they were looking through a window, demonstrated errorless retrieval on the task (Troseth & DeLoache, 1998). These results beget various implications for the nature of transfer, namely, the ability of infants to represent three-dimensional objects and actions on a two-dimensional plane. Furthermore, this last finding suggests that the video deficit effect cannot, therefore, be due only to perceptual factors, and that understanding the nature of the medium is critical to how on-screen stimuli are understood.

Representational Transfer

Hayne et al. (2003) suggest that imitation is age-dependent to reflect an underlying development of representational transfer. This idea is not unique to television, having been established with previous work with scale models, maps, and photographs; this was termed the dual representation hypothesis (DeLoache, 1987). Dual representation discusses the ability to symbolically transfer information between referents and media, specifically stating that if the appearance of the representational object captures the attention of the infant, he or she will have significant difficulty attributing the object to its referent (DeLoache, 1987; Troseth & DeLoache, 1998; Troseth, 2003b).

Off of this theory, Suddendorf (2003) found that 30-month-old infants were better able to transfer symbolic information (i.e. the two-dimensional object on the image represents the three-dimensional object in real life) in an object search task from a photograph than from a video source (69% to 50% errorless retrieval, respectively). While the ability to symbolically transfer information is honed rapidly with age (approaching ceiling effects for both tasks by 42 months), it is clear that children have an initial hurdle to overcome in order to understand secondary representations of real events more fully. The nature of the ‘dual’ representation hypothesis stems from the fact that at a young age, children must be able to supersede perceptual and conceptual disparities between a medium and its referent to understand the symbolic entity itself, as well as the relationship it has to the referent (DeLoache, 1987; Troseth, 2003a; Hayne et al., 2003). This is due to a number of perceptual, cognitive, and social reasons.

Conceptually, as evidenced by examples such as the 'video-window' result (Troseth & DeLoache, 1998), it is clear that the notion of the television as a representational medium is still developing by 24 months. A few established theories fall into three different perspectives assessing why this may be: the perceptual lens, the experiential lens, and the social interaction lens. The first centers on the perceptual impoverishment hypothesis (Schmitt & Anderson, 2002; Barr, Muentener, Garcia, Fujimoto, & Chávez, 2007), which stresses that video stimuli lack the richness of perceptual cues available in the real world; when fewer cues are able to be encoded, comprehension is weaker. When an infant views the screen, an image must reduce all three-dimensional = binocular cues available in the real world to a two-dimensional space within 6° to 7° degrees in the infant's visual field (Anderson & Hanson, 2010). Barr et al. (2007) notes how encoded memory from a televised source is not as visually salient when recalled, resulting in worse performance from video sources, especially for deferred imitation paradigms. Findings such as Suddendorf (2003), Barr and Hayne (1996), and Troseth and DeLoache (1998) can all be explained by perceptual impoverishment.

The experiential position goes beyond the simplicity of the viewer-stimulus interaction because both experience with the medium, as well as overall real-world experience, is lacking in infants. A study by Flavell, Flavell, Green and Korfmacher (1990) assessed understanding of the nature of the medium by inquiring hypothetical scenarios about the objects on the screen. Only 16% of 3-year-olds were able to correctly assert that if the television were turned upside down, the popcorn bowl depicted on the screen would not spill, though this number had increased to 89% by age 5; furthermore, when shown a mascot, 60% of three-year-olds reported that it could ‘hear them’ when they spoke to it (Flavell et al., 1990). These results illustrate how learning and experience with both television and real-world stimuli are the only consistent measures to understand television as a medium separate from reality. Troseth (2003a) designed a ‘training’ paradigm, where the infant was exposed to a video feed of the person hiding the target toy in the adjacent room while the door to that room was open (and hence the child could also witness the same event live in parallel); when ‘trained’ four times, the ability of the 2 year old sample to find the toy rose from 44% to 91% errorless retrievals, demonstrating the influence of even short-term experience on symbolic transfer.

Another reason why transfer is often difficult (which, inadvertently, supports Anderson et al.’s (1981) active model of viewership) is due to the infant tendency to interact with the medium. One of the most fundamental discrepancies between television and the real world is that the latter can respond to the infant’s actions. When infants do not receive such scaffolding and feedback, they do not process information as thoroughly (Pierroutsakos & Troseth, 2003; Mumme & Fernald, 2003; Courage and Setliff, 2010). Social interaction is commonplace in media viewership; as noted by Pierroutsakos & Troseth (2003), 9-months olds attempted to grasp and hit the screen as the stimulus transitioned from stationary to moving; 14- and 15-month-olds exhibited similar frequency of behavior, but were beginning to transition from hitting and grasping to pointing and vocalizing. In the real world, this clear desire to interact would have easily elicited feedback from a live presenter or caregiver. Lastly, Mumme and Fernald (2003) took this idea of social interaction on screen and applied it to observing how emotional information could be transferred from the screen to the viewer. In the study, 10- and 12-month olds witness emotional responses from the on-screen presenter toys; the 10-month-olds were able to infer emotional responses from the televised presentation, demonstrating an aversion to the all toys relative to baseline when one toy was responded to negatively. In contrast, 12-month-olds specifically targeted the negatively treated toy and avoided just that one. Thus, Mumme and Fernald (2003) suggested that social inference from television is possible at a young age, but the transition from generalization to referential specificity, the ability to localize the source to which the emotion is directed, develops later on.

The nature of the medium as only a 'quasi-social' forum for interaction (Strouse & Troseth, 2008), and the inability to effectively assess secondary representations make learning from TV a difficult process for infants. Hayne et al. (2003) asserts how the farther this discrepancy between the medium and the world, the harder it will be for the infant to successfully transfer information. Given these perceptual and cognitive disparities, formal features, which often make depictions even less realistic, have a profound role to play in attention and comprehension.

Imposition of Formal Features

Given the video deficit effect and dual representation, it is critical to observe the effects that formal features have on televised learning, and whether it is exacerbated, ameliorated, or enhanced by them. The diversity of appearance and function of formal features makes this assessment difficult, but overall, features can be divided into subcategories based on their nature (visual or audio), prevalence, and use.

Overall blanket theories on the effects of formal features on attention and learning are not fully established. Nonetheless, the stimulus sampling model (Huston & Wright, 1983) has been adapted and applied to specifically target formal features, termed the ‘feature sampling’ model (Courage & Setliff, 2010). In this light, formal features are not for embellishment; rather, they serve a vital role in the attention and comprehension process (whether intended or not) as markers for important information. Schmitt et al. (1999) summarize the intent of formal features as playing a syntactic role, while content per se forms the semantic; televised conventions such as cuts, audiovisual effects, and pacing changes project to the viewer that a change in attention is necessary to better understand content. Still, Beentjes et al. (2001) concedes that children are more likely to understand formal features that parallel their level of experience. In tandem with Valkenburg and Vroone’s moderate discrepancy hypothesis (2004), the infant viewer must be presented with formal features which are just salient enough to grab attention and highlight important content, but not so complex that they exceed the child’s cognitive abilities. As a result, most studies in formal features have determined that they are often misused by the viewer *and* the producer, and while attention may increase, comprehension is often negatively affected.

The younger children are, the more they are attracted to salient formal features overall (Lemish, 2007). Multiple studies have found that children under two are most attracted to the following formal features: female voices, non-human voices, child voices, animation, and music (Bickham et al., 2001). The same study concluded that young children attend most to rapid pacing than slower pacing Anderson and Levin (1976) found that children between 12 and 24 months pay closest attention to women, children, eye contact, puppets, unusual voices, animations, music, rhyming, and auditory change.

Visual Formal Features

Schmitt et al. (1999) conducted a longitudinal study that demonstrated a significant decline of attention to animation (referring to drawn or computer-generated visuals), with two year olds looking the longest o any group. A study by Greer, Potts, Wright, and Huston (1982) recorded high attention to highly salient formal features, including rapid action, frequent scene changes, and fast cuts. Susman (1978) conducted an attention study concluding that three year olds paid less attention to a prosocial learning program when zoom in effects were added than when a still camera was used. The study concluded that zooms were a distraction from the content (Susman, 1978); since 3-year-olds don't experience zoom effects in real life, and have not watched enough television to be used to them yet, moderate discrepancy says the effects will deter them from understanding.

A marketing research study with three year olds by Macklin (1996) showed that using single-block colors can enhance memory on object recall tasks that share those same colors. Pace and continuity were manipulated in another study by Wright et al. (1984), which found that high paced programs were recalled less accurately in a picture matching task from the program with 4 year olds. Furthermore, low-continuity programs (where the story was parsed in multiple places) also yield a negative effect on the picture task, a measure of memory and comprehension of the narrative (Wright et al., 1984). The authors concluded that, while high paced and high continuity programs elicited more attention (as with Greer et al., 1982) rapid pacing and low continuity negatively affect memory and comprehension; this can be extrapolated to suggest that young children require fluidity to understand program content, and formal features which disrupt this fluidity are cognitively taxing. Strouse and Troseth (2008) encountered a similar result with 24 month olds, finding that imitation scores decreased when a televised imitation paradigm with a toy assembly task was reduced with visual cuts to the middle of each step. Finally, Beentjes et al. (2001) observed unusual camera techniques, one which utilized cuts and dissolves to denote a dream sequence, and another 'shaky camera' to represent a home video. Kids as old as four still were not able to verbalize the intent of the shaky camera until a prologue explaining it was played; dissolves, despite their relatively slower speed, being more 'unconventional' made a assessment of the dream sequence more difficult in four year olds as well (Beentjes et al., 2001).

Auditory Formal Features

Comparatively, studies about audio effects have focused less on attention and more on comprehension, though audio effects are as diverse as their visual counterparts. One study on attention to voices that was run with very young infants 3 to 9 months old (Turnure, 1970) showed that moderate distortions of the mother's voice on a tape recording resulted in equal quieting of motor behaviors, a measure of attention in the youngest babies, and these measures only increased with age.

Interestingly, many studies with voice-overs have been shown to facilitate comprehension. Barr and Wyss (2008) found that voice-overs that included verbal labels for the objects being assembled in a complex imitation paradigm facilitated deferred imitation in 2-year-olds; furthermore, equal facilitation when the voice-over was narrated by a parent or by a stranger (Barr & Wyss, 2008) was observed, citing the effectiveness of the voice-over in itself, not the identity behind it. Seehagen and Herbert's (2009) work yielded a similar result, finding that familiarity of the visible presenter did nothing to influence imitation of a sequence of actions with a toy, but the imposition of a narrative by the mother did so significantly. Neeley and Shumann (2004) observed how using animated spokes-characters can have a positive influence on attention, recall, and even attitudes towards the product being advertised. The second study in this paper revealed that utilizing two 'character' voices, modified in pitch and timbre to "represent the highest level of auditory complexity" (p. 16) found stronger product recall in an animated segment when presented to 2- and 3-year-olds (Neeley and Shumann, 2004) than silent controls. If anything, these findings support that as a formal feature, auditory voice-over (relative to no verbal input) may be one of the few that is beneficial. Barr and Wyss (2008) postulated that this is due to the voice-over reflecting a more naturally occurring phenomenon.

Sound effects, in contrast, are very uncommon in real life; three audio effects were found to elicit higher attention and maintain it an paradigm observing 2-year-olds' attention to television programs: higher volume, special effects (e.g. booms, bangs), and laughter (Gunter and McAleer, 1990). Barr et al. (2009) found that in when a range of sound effects (e.g. swooshes, bells) were mismatched to sequential actions (e.g. when the puppet was put on, shaken, and removed), imitation performance was worse than when they were matched. This finding was only distinct in 12- and 18-month olds— 6-month-olds actually had equally significant imitation between the matched and mismatched conditions, above the baseline and live matched controls (Barr et al., 2009). Barr et al. (2009) asserts that by 12 months, the infant is already utilizing feature signaling, and thus, a logically timed sound effect enhances the information presented. Matched sound effects were actually inhibitory for all age groups when imposed on a live demonstration, supporting that young infants are beginning to comprehend conventions unique to television are nonsensical in the real world (Barr et al., 2009).

Music, perhaps the most frequently used formal feature in children's media (especially infant directed product), also shows very interesting trends in attention and learning. Gunter and McAleer (1990), Bickham et al. (2001) and Schmitt et al. (1999) noted higher attention in 2-year-olds to lively and rhythmic music. A meta-analysis by Bryant, Brown, Silverberg, and Elliot (1981) also recorded particularly high attention scores in infants to faster and exciting tempos, though this attention was brief and declined as the music went on. Barr and Linebarger (2008) and Barr et al. (2010b) stressed that despite music's ability to grab attention, if used incoherently, it substantially negatively influenced learning from video. The first of these studies found that looking time to all televised conditions was high (91%); nonetheless, when a background music track was superimposed onto the demonstration tasks (the same puppet task in Barr et al., 2009), the video deficit effect was exacerbated considerably for all age groups (6, 12, and 18 months), particularly when the music was not played again at the time of test. This indicates that while music may not have much of an effect in a live presentation, it becomes influential when presented as a formal feature in a televised paradigm. Barr et al. (2010b) replicated Barr and Linebarger (2008) with a deferred paradigm (24 hours later) and found the similar result of music enhancing attention but impairing comprehension. These two studies concluded that music, if overused, may create additional cognitive load and distract from content. In Experiment 2 of Barr et al. (2010), imposition of matched sound effects (shown to be beneficial in Barr et al., 2009) while the music was playing, exacerbated the negative effect of the music, but only restored performance to the standard video deficit level relative to live controls.

Other Formal Features

While the majority of formal features fall into the audiovisual spectrum, there have been a smaller number of studies completed on miscellaneous conventions of television viewership, such as repetition and social interaction. Barr et al., (2007), emphasizes learning from repeated exposure to the same stimuli, finding that 12- to 21-month-olds were able to supersede the video deficit after being showed the video stimulus twice as many times as the live demonstration. Conceptually, this may be because in repeating content, infants are given more chances to encode the formal cues of the program into memory, in line with perceptual impoverishment (Barr et al., 2007).

Troseth et al. (2006) ran a study on social interaction with television using three conditions: live presentation, pre-recorded video presentation, and closed circuit video interaction (where the presenter could respond to the infant in real time). Closed circuit interactions led to comparably high performance on an object search task to the live demonstration. Troseth et al. (2006) cited that this was likely due to the presenter's responses to the baby's social cues, such as laughing, pointing, and looking, and wait until the baby was attending to demonstrate. In theory, this finding puts less emphasis on the perceptual limitations of television, and more on the social ones, finding that, as with Troseth and DeLoache's (1998) 'video window' paradigm, the screen itself is not as much a deterrent as is the limitation it imposes on interactivity. Thus, while not inherently 'formal' in the audiovisual sense, these aspects of viewership further the idea that television is a unique medium with its own conventions separate from real life. Overall, screen conventions and formal features need to be learned and understood consciously for television learning to be as strong as live presentation—a feat only achieved with media experience (Anderson & Hanson, 2010). When formal features are designed to fulfill the infants' needs more readily, by logically emphasizing content and allowing for true interactions, etc. they can be beneficial. Otherwise, the video deficit reigns, and, for perceptual, social, and experiential reasons, infants do not learn information as effectively

Thus, high attention to salience alone is not a strong predictor of comprehension, as formal features that elicit attention do not always facilitate comprehension. This suggests that the feature sampling model (Huston & Wright, 1983; Courage and Setliff, 2010) may have a theoretical basis, but unless it falls within the domain of moderate discrepancy (Valkenburg & Vroone, 2004), formal features will most likely serve to detract from comprehension of video sources (with a few exceptions), though attention may be unaffected or even enhanced by them.

The Current Study

Therefore, the aim of the current study was to impose formal feature conditions on a video demonstration paradigm to determine their effects on attention and comprehension. The forced choice task incorporates both direct imitation (McCall et al., 1977; Hayne et al., 2003; Barr & Wyss, 2008) and object search paradigms (Troseth & DeLoache, 1998; Troseth et al., 2006), as the infant must identify a target toy from by imitating the depicted behavior.

Thus, the current study is one of the first to impose range of formal features over the same video source and compare their effects on attention and comprehension relative to controls and each other. Given the research and theories discussed, the following three hypotheses are expected:

- H1. Children will attend to formal feature condition trials more so than control trials, given the infant attraction to perceptual salience
- H2. Children will perform worse on the task on formal feature condition trials than on control trials, given the perceptual, social, and experiential variables associated with comprehension deficits
- H3. Children will perform differently on experimental conditions relative to each other, based on the diversity of findings within each category (e.g. visual, audio, other)

The participants observed in this study were aged 17 to 19 months; this group was selected based on previous imitation research (McCall et al., 1977; Hayne et al., 2003; Barr et al., 2010b; others) suggesting their competency with imitation, but relevant susceptibility to the video deficit. Researcher experience with this age group contributed to this decision; preliminary testing with earlier versions of the task demonstrated relatively little imitation with 14-month-olds, and ceiling effects with 24-month-olds.

Preliminary Feature Analysis

In order to select which formal features to test as experimental conditions, a preliminary formal feature analysis was run on nine commercially produced infant learning videos and DVDs from five different brands (See *Appendix A.1* for sample information), all of which were intended for children 2 years and under. While this analysis was not designed to test the effects of any particular brand, this analysis informed the selection of formal feature conditions used in the main study.

Method

The content analysis was designed to measure relative frequencies (how many times a feature occurred) of formal features in each video. Due to the episodic qualities of these films (short scenes of instruction rather than one continuous narrative), content was analyzed in 60 second intervals spaced 4 minutes apart. Content was analyzed only for the first 30 minutes for each video.

Variables were divided into six categories: voice-over, music, sound effects, visual effects, pacing changes, and camera changes. In some instances, variables were operationalized into a discrete range, particularly in observations of frequency, speed, and volume (e.g. with three options: *throughout*, *interspersed*, and *not at all* for frequency, and *faster/louder than normal*, *about normal*, and *slower/softer than normal* for speed and volume, respectively). In others, variables were entirely discrete, in observing the different types of occurrence within one category (e.g. tone of music, type of visual effect). Only the amount of music category was recoded as an absolute value (in number of seconds). The coding rubric and operationalization of these variable categories can be found in *Appendix A.2* and *A.3*, respectively.

For each 60-second portion, if the degree or type of formal feature occurred at least once, a tick mark was counted towards total frequency, a proportion relative to the entire sample. If more than one category for a given variable was present in a clip, both those categories were tallied.

Results

Findings from the content analysis were not formally analyzed for significance. The following frequencies were the most common throughout the scored sample, by category:

- *Voice-over*: a voice-over was present in 69.84% of clips. Of these, the top three speakers were adult females (46.03%), cartoon voices (19.05%) and child females (15.87%).
- *Music*: music, in any form, was present in 57.14% of clips. Of these clips, the average amount of music was 36.01 seconds, or 57.16% of the entire clip. Almost half of the music heard on screen was instrumental (46.03%), and 54.05% of clips contained a consistent, uninterrupted presentation of music as a background track. Mid-tempo (57.89%) and mid-volume music (77.78%) was most common. Analyses of tone of music demonstrated ‘happy’ (36.51%) and ‘gentle’ (25.40%) tones as the most apparent.
- *Sound Effects*: three most common sound effects used were 1) chimes and bells (34.92%), 2) horns and honks (25.40%), and 3) shakes and rattles

(20.63%), with the most prominent tones being goofy or silly (26.98%) and gentle (25.40%).

- *Visual Effects*: 50.97% of visual effects were presented in interspersed segments. Formal animation was not present in 53.97% of clips. The top three visual effects used were 1) shapes on screen (33.33%), 2) flying or moving objects (20.18%), and 3) distortions (16.67%). The overall color salience of these effects was mostly normal coloring (50.79%), followed by intense, bright coloring (31.57%).
- *Pacing Changes*: pacing for each clip did not stray from normal about half the time (47.62%), though use of slow motion was as twice as common as use of fast motion (23.81% and 12.70%, respectively). This moment was also mostly normal (36.51%), but artificially modified pacing was most common in the jerky or erratic category (20.63%) over the fluid and energized ones.
- *Camera Changes*: camera changes were not very common in this sample of infant learning videos, with 58.73% and 36.51% (for the zooming and angle changes categories, respectively) yielding a stable camera for the entire clip

These findings, as well as previous research, were seminal in determining the test conditions for the main study; thus, the following four variables (two auditory, two visual) were used. They were:

- *Cartoon Voice*: the second most frequent voice-over type, a high pitched, silly-toned voice which was decidedly more salient than the most common voice-over type (female voice),

- *Music*: a background instrumental music track, fast-tempo and happy in tone, that played throughout the entirety of the demonstration,
- *Visual Effects*: an interspersed series of visual 'transition' effects, that included moving shapes, patterns, and distortions to transition to the next frame, and
- *Artificial Dynamics*: a division of the clip into an array of different segments, some of which were slowed down, others sped up, and others frozen, to 1) create a sense of jerky and erratic movement found to be most common in the content analysis, and 2) maintain the original length of the clip.

While some relevant formal features from the content analysis were not included in the experimental conditions, they were otherwise incorporated into the paradigm: for instance, a chime was used to signal the beginning of each clip, and only a still camera (no camera angles or zooms) was used in the screen demonstration.

Method

Participants

Participants were 24 infants (9 Male, 15 Female) from the Greater Boston Area. Ages ranged from 17 to 19 months (522 to 610 days old), with an average age of 18.35 months (550.37 days). Of this sample, 87.5% were white or Caucasian, and 12.5% were Asian American. One additional baby was tested, but not counted due to the inability to finish watching the full demonstration. Participants were recruited via direct mail from birth records obtained at local town halls from the nearby community.

Materials

Lab Set-Up

Infants were seated on a high chair (38 cm L x 41 cm W x 34 cm H) that was attached to a table (61 cm L x 107 cm W x 75 cm H). The participant's caregiver was seated to the left of the participant, and asked not to interfere with testing. Of the sample, 9 participants (37.50%) were uncomfortable with being seated in the high chair, so it was removed and the participants were allowed to sit on their caregiver's lap.

An external monitor display used to present video stimuli (57 cm diagonal) with two sound speakers was placed on a separate wheeled table (46 cm L x 77 cm W x 68.5 cm H). The screen sat 73 cm away from the participant. A child's play pen partition screen (3.5 cm L x 243 cm W x 148 cm H) was set 113 cm behind the display, and covered in a neutral black cloth to eliminate distractions. A video camera (*Canon Optura 200MC*) was mounted 99.5 cm from the table to capture each participant's interaction with the toys, as well as the mirror to reflect the content of the screen.

Infants were presented with live stimuli using one of two specially constructed trays. Trays measured 35.5 cm L x 56 cm W x 4.5 cm H and were made from sanded wood board; two 35.5cm long parallel wooden slabs were glued to the underside of each tray 36.5 cm apart to fit snug between the two arms of the high chair. To fit particular stimuli, one of the trays was outfitted with three 5 cm by 5 cm Velcro patches 15 cm apart from each other. A small, circular bike mirror (8 cm diameter) was mounted behind the table; the mirror was taped to the wall 104.5 cm feet from the floor and positioned to reflect the content of the screen when viewed by the video camera.

Trial Design

The six trials presented to the participants were divided into two types: WHERE trials and WHICH trials. The WHERE trials were composed of three different sets of three similar toys each. All the toys were capable of concealing a smaller plastic animal toy. The WHICH trials were also composed of three different sets of three similar toys each. All the toys could potentially yield a specific effect when activated by the participant. For each set of three WHICH toys, one could be ‘rigged’ such that it would perform the desired effect, and the two would not. It was not visibly obvious from just looking at each WHICH toy if it would work or not.

Toys

Toys used during the warm-up trials were (1) two simple plastic blocks, (2) a transparent cylindrical rattle, (3) a red and yellow toy car that moved its eyes when rolled, and (4) a battery-powered ball that shook when a button was pressed. Toys used during the WHERE trials came in three varieties (see *Appendix E.1* for toy images):

- WHERE 1 was comprised of three concealment toys and one concealed toy:
    - A red plastic funnel with a black top (11 cm H x 10 D)
    - A blue plastic cylindrical cup topped with a yellow wooden cylinder as a handle (7.5 cm H x 9 cm D)
    - A brown plastic teacup turned upside-down with an orange wooden ball as a handle (10.5 cm H x 8.5 cm D)
    - The concealed toy was a small yellow plastic dog (5 cm L x 3 cm W x 5 cm H)

- WHERE 2 was comprised of three concealment toys and one concealed toy:
    - A green plastic cylindrical cup with a white, mushroom shaped handle on top (9 cm H x 8 cm D)
    - A yellow cylindrical cup with a curved top and a red wooden ball as a handle (12 cm H x 8 cm D)
    - A half of a hollow purple plastic container shaped like a bunch of grapes with a green top (8.5 cm H x 6.5 cm D)
    - The concealed toy was a small orange plastic cat (5 cm L x 3 cm W x 5 cm H)
- WHERE 3 was comprised of three concealment toys and one concealed toy:
    - A gray plastic container with a dark blue wooden handle (10.5 cm H x 9.5 cm D)
    - A yellow plastic cup with a purple plastic handle (15 cm H x 7 cm D)
    - A light blue box with a green wooden block as a handle (7 cm L x 7 cm W x 9.5 cm H)
    - The concealed toy was a small green plastic parrot with red and yellow stripes (5 cm L x 3 cm W x 5 cm H)

All concealment toys were presented upside-down with the handles facing upwards, and were each able to completely cover the concealed toy when it was placed underneath. When presented, each WHERE toy was placed on one of three evenly spaced points on the tray, all 9.5 cm away from the child and 15 cm away from each other.

Toys used during the WHICH trials (explained above) came in three sets, and were all meant to produce a desired effect when correctly prompted by the user (see *Appendix E.1* for toy images):

- WHICH 4 was comprised of three hollow toys which could be each rigged to rattle by inserting a collection of plastic beads:
    - A red plastic barrel (8.5 cm H x 7 cm D)
    - A blue plastic bubble solution container with a white cap (9 cm H x 5 cm D)
    - A yellow plastic clay container with a yellow foam handle (11 cm H x 7 cm D)
- WHICH 5 was comprised of three plush dolls that sat in three clear plastic containers (8.5 cm H x 11 cm D); the dolls had the capacity to make noise when shaken or be silent
    - A light blue electronic Cookie Monster doll that grunted when shook and could be turned ON or OFF (8.5 cm L x 21 cm W x 21.5 cm H)
    - A light pink puppy doll in which a noise stick could be inserted (10.5 cm L x 19 cm W x 22 cm H)
    - A light green giraffe doll in which a plastic egg with beads could be inserted (9 cm L x 20.5 cm W x 23 cm H)
- WHICH 6 was comprised of three push-puppet giraffes (9 cm L x 20.5 cm W x 18.5 cm H each), each placed atop of one of three red wooden cylindrical bases (4.5 cm H x 6.5 cm D), To make the toys easier to press, a strip of balsa

wood was glued between the legs of each to be pressed down on with the fingers. The individual differences between each toy were:

- A green and red-necked toy
- A orange and yellow-necked toy
- A pink and green-necked toy

Video Stimuli

Videos were filmed on 16 October 2010. The clips were universally designed to be a still camera shot with only the torso and arms of the experimenter shown, simulating live interaction paradigms used in many studies, where the presenter sat across from the baby. The same basic gray hooded sweatshirt was worn by the experimenter for each video to reduce distractions. The experimenter sat behind the tray with the toys on it, and reached at the toys from behind to present to the camera.

Each specific film clip was between 38.0 and 64.0 seconds long, with an average of 52.40 seconds.. Within each of the six toy sets (three WHERE, three WHICH), the following variables were counterbalanced: (1) the identity of the 'chosen' toy out of each group of three and (2) the position of the 'chosen' toy as either in the center or on one of the sides. Together, both variables form a 2 x 3 matrix of possibilities for each of six trials, forming 36 unmodified videos.

Standardized variables were (1) that the experimenter always used the right hand to lift the concealment toys and the left hand to place the concealed toy in the WHERE trials, the right hand to manipulate the toys in WHICH 4 and WHICH 5, and then both hands to compress the toys in WHICH 6, and (2) that the sequence of events in all trials occurred from right to left (left to right from the baby's perspective). Since the arrangement the three objects switched for each trial, the first object touched was always varied with this arrangement between participants.

Video Demonstrations (see *Appendix E.2 for images*)

The WHERE 1, WHERE 2, and WHERE 3 versions followed this general sequence of showing and hiding from the first fade-in from black:

1. The demonstrator greeted the participant and cued him or her to observe with the phrase "Watch me, look at this!"
2. Using the right hand and moving from right to the left the demonstrator lifted each toy at a 45 degree angle and held it up for 100 ms to show that nothing was underneath it
3. The demonstrator uttered a reaction of surprise ("Huh!") to emphasize the absence of a toy underneath each concealed toy
4. Using the left hand, the demonstrator lifted the concealed toy, held it at the center for the frame for approximately 100 ms, and said "Watch!"
5. The chosen concealment toy was lifted with the left hand and both were held up for another 100 ms
6. The concealed toy was placed at the center of the chosen toy's spot, and immediately covered by the chosen toy in a slow, rhythmic manner so the participant could clearly witness the action
7. The participant's attention was reestablished by the demonstrator mentioning "Watch me again, look at this!"

8. Steps 2-3 were repeated, with the exception of the concealed toy being visible for 100 ms when the chosen concealment toy was lifted; thus, all toys received balanced attention
9. After the second round of lifts, a fade out to black concluded the video clip

The WHICH 4, WHICH 5, and WHICH 6 versions followed this general sequence of events from the first fade-in from black:

1. The demonstrator greeted the participant and cued him or her to observe with the phrase "Watch me, look at this!"
2. Using the right hand and moving from the toy on the right to the left, the demonstrator lifted the first toy at a 45 degree angle, moved it towards the center of the screen, and shook it three times in quick succession to demonstrate its intended effect (for the chosen toy) or lack thereof
3. After a 100 ms pause at the center of the screen, the same toy was shaken three times in quick succession once more to repeat the intended effect (for the chosen toy) or lack thereof
4. An audible verbal reaction was uttered in the vein of "Wow, look at that!" for every toy after each demonstration, for two reasons: (1) not to bias the participant to the chosen toy, and (2) demonstrate an acknowledgement that the cause (and the effect, in one out of three) occurred
5. The participant was instructed to look before each toy was picked up, so as to not bias the child to the first toy, by using a simple exclamation in the vein of "Look!" or "One more time!"

6. Steps 2-5 were repeated for the next two toys
7. Steps 2-6 was repeated for all three toys, so the entire set cycled through twice
8. After the second round, a fade out to black concluded the video clip

Minor exceptions to this filmed procedure were, for WHICH 5, that both hands were sometimes used to place the dolls back into the container if necessary. For WHICH 6, two fingers were used on both hands to demonstrate the compression technique on the push puppets, one hand on each side.

Formal Feature Conditions

All video conditions were designed and edited using *Windows Movie Maker* (v. 5.1). There was one control for each video category, a Standard WHERE condition (SR), and a Standard WHICH condition (SH), which were unedited except for 200 ms fade-ins and fade-outs from black at the beginning and the end, respectively. The experimental conditions are described below.

The first audio condition was the Cartoon Voice condition (CV), where the original audio track of the verbal instruction was muted, and superimposed with a re-recorded script of the same instruction, only with stylistic embellishments designed to make the material more engaging and whimsical, in the manner many cartoon voices present in children's media today (see *Appendix B.1* for voice-over transcript). Once the audio track was re-recorded, it was input into *Audacity* (v. 1.3), a free voice modulation software. The audio track was changed in pitch from E to B♭ (8 semitones) and had its frequency and boost levels altered to 200 Hz and 12 db, respectively.

The second audio condition was known as the Music condition (MU), where the original audio track of the verbal instruction was maintained. A continuous background track of classical music from the infant video *Baby Bach* (Walt Disney Company) was imposed onto each video. The audio clip was located and downloaded from YouTube on 18 October 2010 (URL: http://www.youtube.com/watch?v=4vCeQdeB1pU) The clip was selected for its upbeat tone and medium-quick tempo, and for generally being estimated to be archetypal of the music found in children's programming.

The first visual condition was the Visual Effects condition (EF), which imposed interspersed transitions normally used in children's media to denote a shift in action. Pre-rendered transition effects present in *Windows Movie Maker* (v. 5.1) were used to divide the action. Transition effects were added approximately every 7 seconds to each video. No important actions (such as the hiding of the concealed toy in the WHERE trials) were obscured by the transition. Quick 50 ms fade-ins and fade-outs were positioned directly adjacent to each transition; to make each cut more salient, each video contained 6 to 7 transition effects, depending on the clip's length: (7) 'shatter'; (14) 'checkerboard across'; (21) 'circles'; (28) 'dissolve'; (35) whirlwind'; (42) 'stars'; (49) 'shatter' (see *Appendix B.2* for transition images). Criteria for the chosen transition effects were specifically that they lasted less than 100 ms, took up the entire screen with no bias towards any side or direction (such as ones with panned from left to right), and did not alter the coloring of the video quality in any way.

The second visual condition was Artificial Dynamics (AD), emulated pacing changes common in children's media programs; variations in play speed, including slow motion, fast-forwarded, and paused action, were all utilized to achieve the 'artificial' aesthetic of jerky and erratic movements. As with all other videos, *Windows Movie Maker* (v. 5.1) effects, called 'Slow Down, Half' and 'Speed Up, Double' were used to slow down and speed up the pacing of video, respectively. The voice-over from the control video was re-imposed and matched to the action on screen to maintain normal pitch and rate of speaking. The specific paradigm was designed to both (1) maintain the original video length and (2) specifically be tailored to not obscuring any important task information (see *Appendix B.3*).

In summary, each WHERE and WHICH condition had 6 videos each based on the counterbalanced parameters; each of these 6 videos had 5 edited versions based on the four formal feature conditions (CV, MU, EF, AD) and one control (SR for WHERE and SH for WHICH). Overall, 180 total combinations were possible, but due to between-subject counterbalancing, and the sample size (N = 24), only 144 of these videos were actually made.

Coding System

Each participant was shown a series of 6 videos depicting each distinct toy set exactly once, with each formal feature appearing exactly once, with two controls: one Standard WHERE and one Standard WHICH in a repeated measures design. The individual videos for each participant were combined into one longer video, separated by 10 seconds of black screen in between each. WHERE and WHICH categories were alternated, such that the same task was never demonstrated twice in a row, Secondly, every other participant was presented with a WHERE first alternating with a WHICH first for the adjacent participant, such that between participants, 12 started their presentation of 6 videos with a WHERE and ended with a WHICH, and the other 12 started with a WHICH and ended with a WHERE. Thus, the video versions were alternated both within and between participants. Positioning of toys was counterbalanced between trials such that each participant received three videos where the target toy was in the center and three where the target toy was on either side.

Survey

The caregiver was asked to fill out a survey measuring the infant's prior exposure to media sources. The survey included three ratio questions such as 'How much TV, on average, does your child have per day?', two nominal questions about viewing habits, and eight interval questions with 7-point Likert scales to measure agreement with statements about media consumption, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), with 4 labeled as *neutral*. Of these eight assessments, five were reverse scored, as they expressed a negative statement (e.g. "I believe that children shouldn't be exposed…") as opposed to a positive statement (e.g. "I think early learning can be beneficial…"). A high score on normally scored question (such as 6) correlated directly with a low score on a reverse-scored question (such as 2). The survey is presented in *Appendix C*.

Procedure

Preliminary testing occurred between 21 September 2010 and 20 October 2010, and final testing occurred between 13 November 2010 and 16 March 2011.

Participants were greeted outside of the Psychology Building at Tufts University and led to the lab area. Upon arrival, participant and their parents settled in the lobby outside of the lab area. While the parent filled out a consent form for their child (see *Appendix E.2*), the experimenters familiarized themselves with the participant.

All direct interaction with the participant was completed by the secondary experimenter, while the primary experimenter sat behind the screen controlling the video output. Participants were first presented with one to three warm-up toys depending on their level of comfort, for an average of $M_{WU} = 70.71$ seconds. This was done to demonstrate to the child that he or she was encouraged to interact with toys.

The participant was directed to observe the monitor while the secondary experimenter went behind the partition. The video was played from the beginning (10 seconds of darkness before the chime indicating the start of the visual), to the conclusion of the first clip. The primary experimenter, able to see what the participant was viewing, prepared live version of the trial just demonstrated. For WHERE trials, the primary experimenter placed the three concealment toys exactly as presented on the monitor, with the concealed toy underneath the chosen toy. For WHICH trials, the primary experimenter configured the toys so that the chosen one worked and the other did not, and placed them on the platform (adhered the Velcro for WHICH 5 and WHICH 6) in the positions shown to the baby. Thus, the toy to the left of the baby was the same toy that was just witnessed on the left side of the screen.

The primary experimenter paused the video in between each clip during the blackout. Within 10 seconds of finishing the clip, the participant was presented with the live action version of the platform shown in the video with the exact same toys on it. The secondary experimenter brought the tray from the right (because the monitor made it impossible to place it directly from the front). In order to eliminate a presentation bias, the secondary experimenter placed the tray on the table roughly 30 cm out of the participant's reach for 200 ms. Once the participant had a chance to see the stimulus, the secondary experimenter pushed the tray directly forward within reach of the participant (see *Appendix E.3* for toy presentation).

For WHERE trials, the secondary experimenter asked the participant, "Can you find the [animal toy]?" inserting "doggy," "kitty," and "birdie," for WHERE 1, WHERE 2, and WHERE 3, respectively. For the WHICH trials, the secondary experimenter asked the participant, "Can you shake the one that makes noise?" for WHICH 4 and WHICH 5, and "Can you press the one that falls down?" for WHICH 6. The secondary experimenter stood right of the participant, with no further interaction. The participant was given 60 seconds to respond to the live stimulus, and encouraged when he or she performed the task (correctly or not).

After 60 seconds, the secondary experimenter removed the tray, the cue for the primary experimenter to press play again. The secondary experimenter returned the tray to the primary experimenter, who cleared it to prepare for the live stimuli for the next video. The entire process was repeated six times until the participant finished interacting with the sixth set of stimuli.

Scoring

Each participant was scored for proportion of looking time (the number of seconds viewing the screen per clip over the total length of the clip), order in which each toy was touched and acted upon using the 'target action,' and latency until said behaviors occurred on each toy. Further details on these scoring variables are noted in the Results. Scoring forms for during- and post-test data are displayed in *Appendix D*.

Results

Because the within-samples design ordained that each participant was exposed to two controls (SR and SH) and four experimental trials (MU, CV, EF, and AD), not every participant received every task paradigm in a given run. While each participant received both controls (N = 24), between subjects each participant received a counterbalanced combination of formal features for the remaining four trials, so for each formal feature condition, n = 12 participants were presented with it imposed on a WHERE trial, and n = 12 on a WHICH.

Preliminary comparisons of all means (looking proportion and target action order) from each n = 12 condition between the WHERE half and the WHICH half yielded non-significant differences, confirming that any subsequent effects were not due to the nature of the WHICH task versus the WHERE task. Thus, data was collapsed across trials, and each condition was calculated as N = 24.

Looking Time

Infants' attention to the clip was examined by comparing the proportion of looking time (percent) to each clip. Looking time was scored on a computer using a push-button technique, where a button held down when the infant looked at the screen, and released when he or she looked away. The mean proportions of looking time for each condition are compared in Figure 1. All conditions had a high proportion of looking at the screen, $M_{looking}$ = 88.28% over all conditions. As Figure 1 shows, the one minor departure was that attention in the AD condition appeared to be lower, but infants still looked at the screen over 75% of the time (M_{AD} = 78.95%).

Figure 1. Percentage of Looking Time During Video

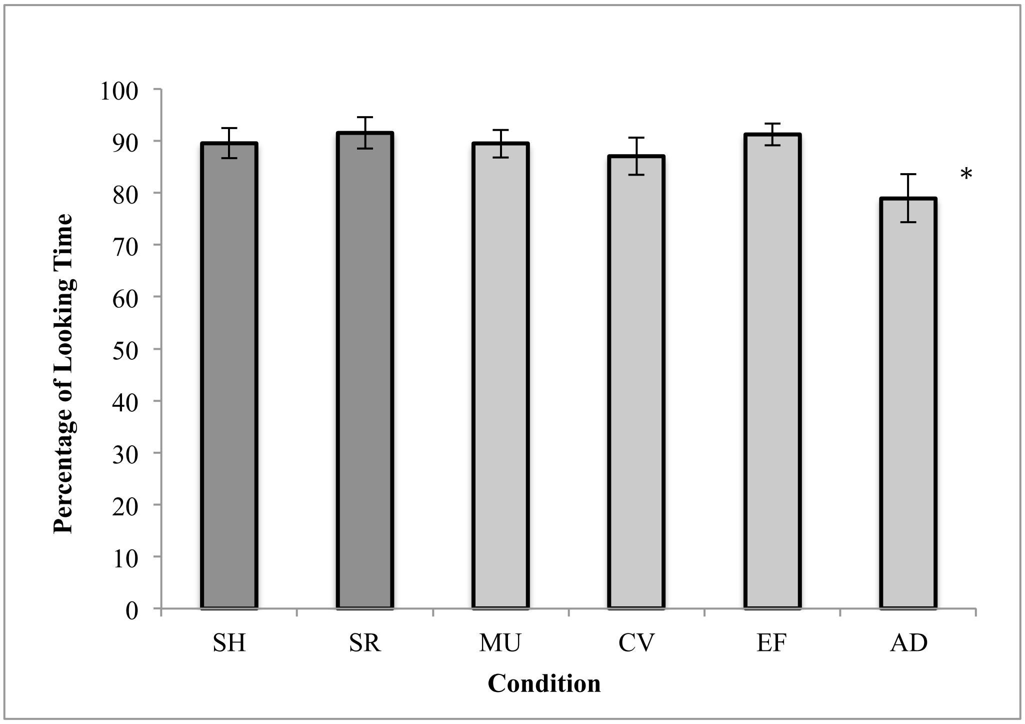


Figure 1. All values are measured in relative proportion of looking time ($M_{looking}$ = 88.28% seconds). Error bars represent +/- standard error from the mean, determined from the overall sample (N = 24). Asterisk denotes statistical significance, $p < .05$. Control trial columns are shaded for clarity.

Two one-way, repeated measures ANOVAs were completed, comparing looking proportion of each of the two controls and the four different formal feature conditions. The comparison between the Standard WHERE control and each of experimental conditions was significant, F (4, 119) = 5.47, $p < .001$; post-hoc analyses revealed that AD was lower than SR ($p < .05$), MU ($p < .05$) and EF ($p < .001$). The ANOVA comparison with the Standard WHICH control was also significant, F (4, 119) = 3.65, p = .0083; AD was lower than SR ($p < .05$), MU ($p < .05$) and EF ($p < .01$).

Another one-way, repeated measures ANOVA with all six samples was run to confirm these results, and again found a significant condition effect F (5, 14) = 3.681, p = .004, with pairwise analyses confirming a significance difference between AD and all other conditions AD < SH (p = .032), AD < SR (p = .001), AD < MU (p = .006) AD < CV (p = .023) and AD < EF (p = .009). No other condition differences were observed.

Participant observation of the two 'critical actions' per clip was also noted. The critical actions were defined as the specific steps of each clip that denoted the core of the task: for WHERE trials, the first was the hiding of the hidden toy and the second was the reveal of the hidden toy after the second lift; for WHICH trials, the critical action was simply both times the 'working' toy was demonstrated. Each participant either looked at both critical actions, one of the two (either), or neither for any given clip.

Table 1 depicts observation of critical action; no significance tests were run due to the nature of the correlated sample, but clearly, both critical actions were viewed most of the time (and at least one almost all the time), as can be expected by such high looking proportion means. Music was the only condition (along with the controls) that had over 66.67% of participants viewing both critical actions. The chance of missing both was negligent, resulting in only two trials overall where neither critical action as observed.

Table 1. Observation of Critical Action

	SH	**SR**	**MU**	**CV**	**EF**	**AD**
Viewed Both	21	20	22	18	18	18
Viewed One	3	3	2	5	6	6
Viewed Neither	0	1	0	1	0	0

Table 1. Each cell denotes how many participants (N = 24) viewed each number of target actions per condition. Each video contained two target actions revealing the core of the target task. Control trials are shaded for clarity.

Target Action Order

Completion of the target action was quantified in an 'imitation score' for each participant, based on the order in which the target action was completed target toy. During scoring, if the participant performed the target action on the target toy first (finding the toy for WHERE and activating the toy for WHICH), they received a score of *1*, if they performed the target action on the target toy after performing the target action on another toy first, they received a score of *2*, and a score of *3* was given if the target action was performed on the other two toys before it was performed on the target toy. If the target action was not completed at all, a score of *4* was given.

To determine each condition's elicited performance relative to chance, a binomial probability was calculated relating each condition to chance, or p = 1/3. Analyses were calculated using the 'first (a score of *1*) vs. not-first (a score of *2*, *3*, or *4*)" dichotomy, comparing the number of times the participant performed the target action on the target toy first relative to chance (baseline) of k = 8, as shown in Figure 2. These analyses yielded that both controls SR (p = 0.028) and SH (p < .0001) were significantly above chance, begetting a higher ratio of *1*s to not. In contrast, every experimental condition except CV (p = 0.017) was found to be at chance level. This demonstrates that, without concluding relationships between conditions, the two controls and the Cartoon Voice condition had a positive affect on learning the task, while the other three experimental conditions (MU, EF, and AD) did not.

Figure 2. Target Action Completed First on Target Toy

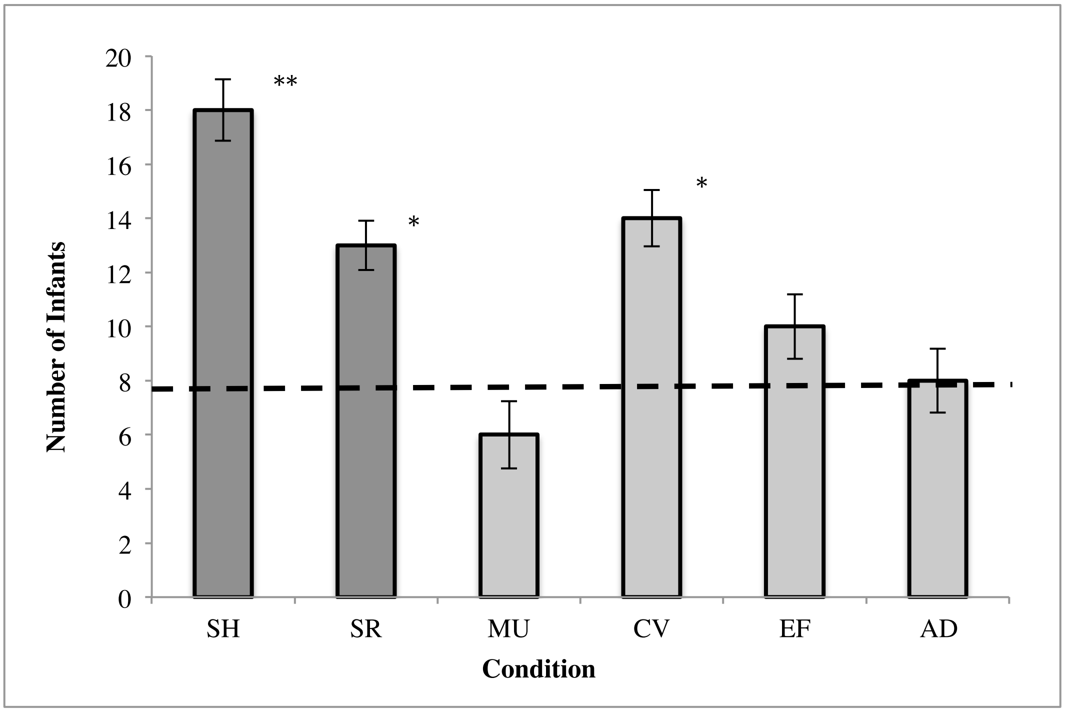


Figure 2. Columns represent the number of infants who performed the target action first on the target toy as a measure of learning. The dashed line marks chance level at $k = 8$, 1/3 of the sample (N = 24). Error bars represent +/- standard error from the mean, determined from the overall sample. One asterisk (*) denotes statistical significance above chance, $p < .05$; two asterisks (**) denote statistical significance above chance, $p < .0001$. Control trials are shaded for clarity.

Comparative target action order analyses were completed using McNemar's Test for Correlated Proportions, comparing the response patterns of each individual participant on a 2 x 2 matrix between any two conditions. McNemar analyses were completed for the 'first vs. not-first' dichotomy; each participant was placed into one of the four quadrants based specifically on his or her pattern of response between the two conditions being considered (e.g. if comparing MU and AD, the top left quadrant would receive one tally if one given participant scored a *1* for both MU and AD, the top right if they scored a *1* for AD but not for MU, the bottom left if they scored a *1* for MU but not for AD, and the bottom right if they did not receive a *1* for either). The quadrants to be compared were the bottom left and top right, as they showed an opposite effect between conditions.

The first major result compared the WHERE (SR) and WHICH (SH) controls (p = .1797) determining no significant difference between the two. Both values were also comparably above chance in the binomial probability analysis. These findings allowed the control conditions to be collapsed into a universal control (S). Rather than comparing each experimental condition to one control over the other, each participant's score on the experimental condition was compared to the control that matched the type of task (either WHERE or WHICH) for that participant; in other words, if the MU condition for Participant 1 fell on a WHERE trial, that score for MU was compared to the SR (the WHERE control) score on the matrix; conversely, if for Participant 2, the MU condition was on a WHICH trial, that score for MU was compared to the SH (the WHICH control) score on the same matrix, with the four quadrants still summing to N = 24. The method made the comparison of conditions to controls stronger because any potential differences between the WHERE and WHICH modalities were controlled for (even though the difference between them was non-significant).

Table 2. McNemar Analyses for Conditions versus the Universal Control

2.1*		S (Universal Control)	
MU (Music)		1st	Not 1st
	1st	4	2
	Not 1st	12	6

2.3		S (Universal Control)	
EF (Visual Effects)		1st	Not 1st
	1st	5	4
	Not 1st	9	6

2.2		S (Universal Control)	
CV (Cartoon Voice)		1st	Not 1st
	1st	8	5
	Not 1st	8	3

2.4*		S (Universal Control)	
AD (Artificial Dynamics)		1st	Not 1st
	1st	6	2
	Not 1st	10	6

Table 2. McNemar analyses compare relative pairings (N = 24 of collapsed sample) for analysis of individual performance between two conditions (each experimental condition versus the universal control, S) from the same participant. These analyses are dichotomized by '1^{st} vs. Not 1^{st}' performance of the target action on the target toy. Shaded quadrants denote opposite performance between trials and are the regions of interest. **Tables 2.1** (Music) and **2.4** (Artificial Dynamics) demonstrated statistically significant condition effects relative to S, denoted by asterisks.

Table 2 shows that McNemar analyses between the universal control S and each experimental condition indicated two significant differences: the Artificial Dynamics condition (Table 2.4) yielded a significantly lower proportion of first to not first than S (p = .0386), as did the Music condition (Table 2.1) relative to S (p = 0.0065). By these same standards, the Cartoon Voice (Table 2.2) did not differ significantly from the universal control S (p = .2905), in line with the binomial probability result. The Visual Effects

(Table 2.3) condition did not differ from the universal control ($p = .1334$), but was found to be significantly lower than each control (SR $p = .0461$; SH $p = .0212$) separately.

Between experimental conditions, the collapsed WHERE/WHICH method used for universal S could not be used, because this resulted in total ns that were too small to yield any significant result; thus, each participant was compared with N = 24. The following listed outline the four significant comparisons of the total six conducted between every pair of experimental conditions (MU, CV, EF, AD):

- CV was greater than AD ($p = .0352$)
- CV was greater than MU ($p = .0215$)
- AD was greater than MU ($p = .0327$)

All other combinations (CV = EF; AD = EF; MU = EF) yielded no significant results.

McNemar analyses were also completed on the inverse dichotomy of 'did (*1*, *2*, and *3*)' vs. 'did not (*4*)', in order compare all infants who eventually learned from the video to those who did not. All McNemar analyses between S and each experimental condition were non-significant, indicating that even with all imposed effects, infants were able to complete the task at hand (ignoring whether they did it on the correct toy first).

In summary, based on these original McNemar finding from the 'did first' versus 'did not first,' dichotomy, comparing conditions to control and the conditions to each other, Table 3 summarizes a rank order for effects, from highest to lowest performance.

Table 3. Summary Table for Condition Effects

Condition	S	MU	CV	EF	AD	**RANK**
S	—	S > MU $p = .0129$*	S = CV $p = .2905$	SR > EF; $p = .0461$* SH > EF; $p = .0106$*	S > MU $p = .0129$*	–
MU	MU < S	—	MU < CV $p = .0215$*	MU = EF $p = .3438$	MU < AD $p = .0327$*	**4**
CV	CV = S	CV > MU	—	CV = EF $p = .1796$	CV > AD $p = .0352$*	**1**
EF	EF < SR EF < SH	EF = MU	EF = CV	—	EF = AD $p = .6875$	**2**
AD	AD < S	AD > MU	AD < CV	AD = EF	—	**3**

Table 3. This chart summarizes the numerous relationships between experimental conditions (Hypothesis 3), with *p* values denoting significance (asterisks denote $p < .05$). The far left column expresses the rank from least detrimental (1) to most detrimental (4) effect on task performance based on these relationships. The shaded and non-shaded areas mirror each other, with each variable switching sides of the equation based on the corresponding columns and rows of the table.

It is clear that in terms of strategic action on the target toy, children performed best in the control conditions with no formal features imposed. The Cartoon Voice condition (CV) performed at the same level as the controls. In contrast, Artificial Dynamics (AD) and Music (MU) yielded the weakest results in terms of performance, both were universally weaker than S and CV; furthermore, MU was weaker than AD, and the weakness of MU relative to CV was more highly significant than the weakness of AD to CV; as a result, MU was considered the weakest condition, and AD the penultimate worst. The Visual Effects (EF) condition was the most inconsistent condition, demonstrating lowered performance relative to control, but exhibiting no statistical differences from any of the experimental conditions; still, due to more robust weaknesses in MU and AD, and overall parallels between CV and the control conditions, EF was judged most aptly to be overall worse than CV and S, but better than AD and MU. The following rank order is presented in Table 3, and is reflected in the relative column heights of Figure 2.

The most significant pattern of results demonstrates that overall, babies exhibited a learning deficit with the imposition of all formal features, except for Cartoon Voice.

Survey

A survey was conducted with the caregiver of each participant (N = 23; one survey was excluded due to partial completion) to gauge media consumption patterns. These variables were not used as criterion for significance tests, but rather to serve as a stronger basis for discussion.

The survey indicated that overall, the infants participating in the study (M_{age} = 18.35 months, or 550.33 days) watched over an hour of television per day (M = 71.83, SD = 76.14) indicating a widely dispersed range of responses from 0 (n = 3) to 270 (n = 1) minutes per day. Only six out of 23 participants (26.09%) reported older siblings M_{age} = 49.67 months) in the household, which often affects viewing time and content in the home. Only four participants (17.39%) recorded using baby videos in the past, and six (26.09%) mentioned subscribing to a TV channel marketed towards babies or infants.

Inquiries included a 7-point Likert scale assessment of parental habits and opinions on infant television viewing, which asked parents to rate their degree of agreement with eight statements. In tallying the eight scores recorded for each participant, an overall agreement score was calculated for each participant. Agreement scores were calculated by using an equation that flipped scores from reverse scored statements and averaged them with normally scored statement scores. Reverse statement scores, rated from *1* to *7*, were subtracted from 8, and then averaged over 8 to obtain each participant's final score. Therefore, the attraction score equation used was: AS_{par} = ((8 –

$st_1) + st_2 + (8 - st_3) + st_4 + (8 - st_5) + (8 - st_6) + st_7 + (8 - st_8)) / 8$, with '$st_1$' referring to statement number 1, and so on. The overall mean attraction score was obtained by summing each participant attraction score and dividing by N = 23.

The overall mean score calculated was $M_{AS} = 3.28$, $SD_{AS} = 0.81$; as a singular sample of 23 parents, it was compared to a hypothetical population mean of 4 (neutral on the 7-point Likert scale), and found to be significantly lower $t\ (22) = 4.2501$, $p < 0.001$, indicating an overall moderate disagreement with infant viewing of televised stimuli.

Discussion

The present study demonstrated that certain formal features have a significant effect on the attention and comprehension of 17 to 19 month olds, finding differences between specific features. Only one formal feature was found to have a significant influence on attention: the Artificial Dynamics (visual) category decreased overall high attention to the screen (Hypothesis 1). The Music (auditory), Visual Effects (visual), and Artificial Dynamics conditions were all at chance level for comprehension (task performance), while the control trials and the Cartoon Voice (auditory) were significantly above chance. Both the Music and Artificial Dynamics conditions yielded significantly lower performance on the target task than the universal control, while the Visual Effects condition was inconsistent, but showed weaker performance than both controls separately (Hypothesis 2). Lastly, between conditions, the Cartoon Voice condition elicited higher performance than the Artificial Dynamics and Music conditions, and Artificial Dynamics was stronger than Music. Therefore, the assessments of the original hypotheses are as follows (Hypothesis 3):

H1. Hypothesis 1 was not supported, because attention to the screen ubiquitously high during the demonstration; only the Artificial Dynamics condition was observed significantly less time than the controls and the other conditions, although it was still high enough to conclude that the depletion in attention was minimal

H2. Hypothesis 2 was supported for almost all conditions; Artificial Dynamics, Music, and Visual Effects all performed at chance level. Viewing the Artificial Dynamics and Music conditions yielded significantly worse performance on the target task relative to the universal control, and Visual Effects yielded significantly worse performance than each control separately. Only the Cartoon Voice condition maintained performance above chance, the same as the controls

H3. Hypothesis 3 was supported, as condition effects were found between three of the four conditions: the Cartoon Voice yielded greater performance than Artificial Dynamics and Music, and Artificial Dynamics yielded stronger performance than Music. These four categories were ranked according to their effect (4 being the most negatively influencing):

1) Cartoon Voice – no effect relative to controls
2) Visual Effects
3) Artificial Dynamics
4) Music

Therefore, the majority of formal features tested in the current study had a negative influence on target task performance, though it should be noted that this is compared to unexpectedly high performance on the baseline conditions. All formal features (except Cartoon Voice) performed significantly below the controls, but at chance level, stressing that these results do not support a prominent video deficit effect in general, despite that they do suggest the negative influence of formal features.

Attention

The target age group was expected to be the most fixated on salient cues relative to older groups (Wartella & Ettema, 1974; Strasburger et al., 2009; Anderson and Hanson, 2010), and clearly, this appeared to the be the case with their high viewership over all conditions. Nonetheless, changing the presumed salience (formal feature conditions were judged to be more highly salient than controls) of the on-screen stimuli did little to affect overall fixation to the screen. Rather, this suggests that on a whole, salience was high to begin with, and infants were engaged in the demonstration for the majority of the time it played. Qualitatively, the most common reasons for looking away were usually because of an explicit distractor (e.g. becoming preoccupied with the mat on the table); otherwise, all conditions except the Artificial Dynamics condition were evidently salient enough to maintain a high level of attention. In comparison to other attention studies, the overall mean looking time ($M_{looking}$ = 88.28%) was higher in comparison to previous work, surpassing Barr et al.'s (2010a) and Seehagen & Herbert's (2009) analyses of attention, who found 70% and 81.83% looking ratios, respectively.

The hard data reflects numerous qualitative observations of infants clearly engaged with the material. These frequent instances of smiling, laughing, and bobbing their head to the music as it played all implicate Anderson et al.'s (1981) active model of attention. The active theory of attention can help explain high viewing time, mainly because it posits that children's desire to interact with the medium keeps them focused on it longer. Also applicable is Huston & Wright's (1983) exploration search model, which could be used to explain that high overall attention resulted from lack of experience with the medium, causing infants to explore the most salient stimuli most thoroughly. Attentional inertia (Anderson et al., 1979) may have played a role in garnering such prolonged attention times. Anderson and Levin (1976) noted that infants will instantly fixate on certain variables such as auditory change. Because each clip was preceded by a chime sound effect to signal the start of relevant content, this may have prompted the impetus for attentional inertia, resulting in no further need to look away unless a more complex stimulus caused a distraction (Wartella & Ettema, 1974).

The longest clip used was 64.0 seconds long, which may have mediated the overall high attention to the video demonstration. Anderson and Levin (1976) and Wartella and Ettema (1974) discuss how infants younger than two cannot remain fixated on one stimulus for more than one minute overall—thus, had the demonstration been longer (as in Barr & Linebarger, 2008, which utilized 90 and 120 second demonstrations), overall attention may have been lower, but still not showing any pronounced differences from the formal features. The control trials in themselves were salient enough to elicit such a response, but different enough in design from the experimental conditions to merit a specific change in looking time.

The Artificial Dynamics condition was the exception as the only condition attended to less significantly than the controls and the other experimental conditions. However, as its own value, it was still considered high (M_{AD} = 78.95%), especially when compared to analogous values in previous research (Anderson et al., 1981; Seehagen & Herbert, 2009). Nonetheless, in order to retain consistent timing relative to the other videos, a very specific pattern of alternating slow and fast motion was used in the Artificial Dynamics condition, which always began with a 20 second period of slow motion; the voice over was unaffected, but the first third of each video was arguably low in salience, before the first critical action even occurred. Anderson and Levin (1976) noted slow pacing as one of the few variables that actually decreased attention; thus, early onset slow motion likely decreased attention to the Artificial Dynamics condition.

Task Performance

Given that Hypothesis 2 was the most strongly supported of the three, it is clear that most formal features had a negative effect on learning from video. Nonetheless, the way in which this effect was manifested was curious; rather than exhibiting results that indicated performance below baseline for formal features, the controls (and the Cartoon Voice condition) performed significantly above baseline, while the remaining formal feature conditions were at chance. For the purposes of this study, performance significantly above chance illustrates that some pattern of learning must have occurred from the stimuli, while at-chance performance maintains that learning did not occur. Nonetheless, if a video deficit were present, it would seem that every column on Figure 2 would have to shift downwards to reflect a presumed difference from a live control. Still, admittedly, with the current design, it would have been difficult to find below-chance performance unless the sample size were substantially larger.

Furthermore, while all significant analyses were conducted on the 'first versus not-first' dichotomy, the 'did versus did not' dichotomy revealed that all conditions exhibited more participants doing the task at all than not. In looking at the number of children who completed the task first for both controls (n = 13 for Standard WHERE and n = 18 for Standard WHICH) and the Cartoon Voice (n = 14) it seems implausible that infants could theoretically do much better overall. Given that an 'ideal' sample would still require room for error, comparing these results to such a sample would likely not yield a significant difference. In other words, though comparing to a live condition in this sample was not part of the original design, this hypothetical comparison loosely refutes a video deficit effect for these tasks.

The video deficit effect is common in a number of video studies that use different measures, particularly imitation (McCall et al., 1977; Barr & Hayne, 1999; Barr et al., 2009), and object search paradigms (Troseth & DeLoache, 1998; Troseth, 2003b). Nevertheless, a number of studies have also concluded that depending on the method of presentation, transfer of information can be comparably robust. Meltzoff (1988), one of the first studies to examine comprehension of video found that infants performed much better on deferred imitation than on the two baseline controls, both of which still used the television: the first baseline condition depicted the demonstrator on screen, without exposing the participant to the target toy or the target action (and thus assessed spontaneous propensity to perform the act); the second showed the demonstrator manipulating the toy, but not in any relevant or consistent way. The clear inability to perform the task in these two conditions showed that when the demonstration was purposeful, the infants were able to learn from the screen, especially the older group, 24 month olds (Meltzoff, 1988).

Mumme and Fernald's (2003) work with emotional information expressed that children as young as 10 months had learned to avoid stimuli that had garnered a negative response on video, and by 12 months, infants could specify the negatively tarnished toy with significant accuracy. Even studies that support the video deficit, such as Barr and Hayne (1999) found that the effect was largely dependent on age— specifically, that it became weaker (the difference between live and video became less significant) as the sample got older, finding the least significant difference with 18-month-olds over 12- and 15-month-olds. In other words, though a video deficit was present relative to live controls, it was evident that children were acquiring proficiency with the medium by 18 months. Barr et al. (2007) emphasized repeated exposure to the demonstration as an ameliorant for the video deficit. Specifically, they found that 12 to 21 month olds (again, with a positive age effect) performed better on the deferred imitation task when the stimulus was repeated twice as many times. Troseth (2003a) also utilized a 'training' paradigm where infants were given the opportunity to witness the target task four more times than the control, also successfully improving object search to 91% errorless retrieval in 24 month olds. This finding in particular demonstrates that grasping of secondary representations can be facilitated with repeated exposure. Barr et al. (2007) attributed this facilitation to having more chances to digest the perceptual cues necessary from the two-dimensional source; according to the perceptual impoverishment theory, the more opportunities children have for symbolic transfer, the better at it they will be.

Thus, if observing the current results through the experiential lens of comprehension (which emphasizes experience with television to learn its conventions), both the relative age (long term) and repeated exposure to the same demands throughout the entire test may have contributed to overall high performance on the task. The current study involved repeating the same target actions over three trials each.

Furthermore, results of the parent survey indicated that on average, the majority of the sample was well experienced with media at the time of test. As Anderson and & Hanson (2010) reiterate, experience with television and the real world will continue to hone skills of representational transfer. Thus, if viewed through the experiential lens, the current results of strong baseline performance are justifiable, despite a non-apparent video deficit. Given the proper opportunities, children can substantially learn from video.

Therefore, the question of interest is how individual formal features, which did demonstrate the video deficit, exacerbated task performance (except for Cartoon Voice, which matched the controls). Given the influential nature of the formal features applied, it is clear that Hypothesis 2 was supported, substantially for Artificial Dynamics and Music and somewhat for Visual Effects.

Cartoon Voice

Cartoon Voice yielded no condition difference with the controls; rather than inciting chance-level performance on the task as the other effects did, it indicated no video deficit, suggesting that it may have been neutral, or even facilitative to completing the task. Originally, Hypothesis 2 predicted that despite condition differences, all formal feature conditions would be worse than the controls; yet, some literature suggests that this condition, which did nothing but modify the quality of the voice-over, helped children comprehend video content. Barr and Wyss' (2008) results can be used to defend the current findings; they found that verbal labels as the demonstration occurred, *regardless* of the familiarity of the speaker (mother versus stranger), significantly increased deferred imitation on the assembly task. Voice-over can override visual information if the auditory quality demonstrates the developmentally appropriate level of salience to facilitate the content presented. Neeley and Shumann (2004) found this to be true when animated character voices, similar in description to the one used in the current study, enhanced both attention and memory in 2-year-olds. Seehagen and Herbert (2009) also promoted the positive influence of the voice-over, finding that narration overlaid onto the video enhanced imitation performance.

The theoretical basis behind the voice-over benefit, particularly when enhanced in pitch and timbre, can be explained by both the feature sampling (Huston & Wright, 1983) and moderate discrepancy models (Valkenburg & Vroone, 2004). Via the latter, if voice-over appears to have a facilitative effect on learning, then this method of instruction fits well within the young infant's (17 to 19 month olds') cognitive capacity, but challenges him or her just enough to remain engaged with the medium. Feature sampling (Huston & Wright, 1983) demonstrates how the increased salience of the voice served to benefit the content, highlighting that it was worth listening to. The voice itself, unlike the normal voice-over present in every other video, was continuous throughout, rather than pausing between critical actions. Wright et al.'s (1984) finding that high continuity in programming facilitated recall further supports high performance on the current task.

Thus, voice-over is a strong positive influence on learning from video, primarily for its clarity and continuity.

Visual Effects

Nevertheless, every other experimental condition yielded the opposite effect, actually inhibiting performance back to baseline and supporting the video deficit. In the ranking established by the ANOVAs and the McNemar analyses, Visual Effects was ranked second, essentially the least detrimental of the three effects that resulted in baseline performance. Given that Visual Effects ranked lower than both the controls, but was curiously equal to all other conditions, it was easily the most inconsistent of all the conditions tested, despite infants' high attention to it overall (M_{EF} = 91.24%).

The most likely explanation for this detriment was that the nature of the effect's presentation resulted in a highly fragmented visual presentation. Visual transition effects were imposed every 7 seconds in this condition, often resulting in what appeared as fragmented 'cuts' from one section to the next. This disruption in continuity can be paralleled to conditions encountered in Wright et al.'s (1984) continuity paradigm. While the published literature utilized produced content, the parsed condition in the current study can also be defined as 'low continuity.' Thus, in line with Wright et al.'s (1984) finding, low-continuity yielded negative effects on memory and comprehension, relative to the fluidity of the 'high continuity' programs (such as the Cartoon Voice condition in the current study). Additionally, Strouse and Troseth (2008) found that imposing cuts in their video demonstration exacerbated the video deficit in a 24 hour deferred paradigm (though their paradigm included shortening exposure to the stimulus).

It is apparent that the visual disruptions caused by such a condition, while still salient enough to maintain an orienting response, did little for comprehension. In context of the feature sampling model (Huston & Wright, 1983), the disjointed presentation was not an effective marker for relevant content. Quite the opposite— frequent transitions could have been misinterpreted as the beginning of a new segment, confusing the infant viewer. It is possible, therefore, that subconscious mental representations of these conventions (i.e. cuts are used to transition to another time) are still developing, and are being disrupted by inappropriate use of formal features (Beentjes et al., 2001). In terms of Barr et al.'s (2007) perceptual impoverishment theory, the frequent cuts of this condition broke the narrative flow so often that infants did not have enough opportunity to encode the information on the screen before a transition effect parsed the action, hindering representational transfer between the stimulus and how it would appear in the real world.

Transition effects, despite their frequent use in infant learning videos (a total of 70.18% utilized shapes, flying objects, and distortions in the preliminary content analysis) are a more 'media-centric' effect, entirely non-existent in the real world (unlike a voice-over, which can be 'borrowed' from the real world). For these reasons, lack of experience pushes them outside of moderate discrepancy (Valkenburg & Vroone, 2004) for the infant age range, resulting in inconsistent performance on the target task.

Artificial Dynamics

Like the Visual Effects condition, the fundamental detriment from the Artificial Dynamics condition also resulted from a fragmented, non-continuous presentation. This condition, as a series of parsed slow motion, fast action, and paused segments, fostered at-chance imitation scores and a significant lowering of attention. In comparison to the Visual Effects condition, the Artificial Dynamics condition is even less regular in segment parsing, relying on apparently arbitrary changes (though in reality, they were controlled to not explicitly obtund the critical action presentation) rather than regular fragmentation ever 7 seconds. Thus, this difference is reflected in the current finding that Artificial Dynamics ranks below Visual Effects in its effects on comprehension.

From a theoretical standpoint, the relatively lower looking time in the Artificial Dynamics condition can be attributed to a decreased interest in lowered salience. Given that this condition began with approximately 20 seconds of slow motion before the first critical action was presented, as Anderson and Levin (1976) noted, slow pacing was one of the few formal feature conditions that lowered attention. According to the exploration search model (Huston & Wright, 1983), because of this depressed level of, infants likely had no motivation to explore the content presented, and thus did not learn as much as they could have. This condition was the only one that loosely suggests a direct relationship between attention and comprehension, for a number of reasons: firstly, according to Barr et al.'s (2007) perceptual impoverishment view, decreased looking due to disinterest and distractibility created less opportunity for the already difficult task of deriving two-dimensional information from three-dimensional. Thus, the reduced comprehension in the Artificial Dynamics condition could have resulted from looking less, increasing the chance of encoding the target action and not knowing how to proceed. This is somewhat unlikely, given that 1) despite significant condition effect in attention, looking time during this condition was still over 75%, and 2) analyses of critical action observation (were they looking during each critical action) showed that only 25% of participants missed one critical action, and none missed both in this condition.

From the experiential lens, while slow pacing has been related to better comprehension than rapid pacing (which, recall, the stimulus also included; Wright et al., 1984), like sound effects and visual distortions, slow motion is not a natural occurrence in the real world. In line with the moderate discrepancy hypothesis (Valkenburg and Vroone, 2004) the farther the visual depiction (slow pace) is from the referent (the action in real time), the less likely the infant will be able to make the association between the action on screen and the same action in real life. By this logic, Christakis and Zimmerman's (2009) and Anderson and Hanson's (2010) ideas on relying on experience (both real world and with the medium) to understand symbolic representation directly bear on whether or not features are used effectively by the viewer. The disjointed realism of Artificial Dynamics, with which participants are not experienced, likely impeded feature sampling (Huston & Wright, 1983), resulting in worse performance on the task.

Music

Lastly, the results implicated the Music condition as the most detrimental to learning from video, exacerbating the video deficit effect despite high levels of attention. Unlike the Visual Effect and Artificial Dynamics conditions, the music used in this condition was a highly continuous background track; thus, the theoretical framework behind music's effects on learning is fundamentally different.

Attention to the Music condition, like the others, was also very high overall (M_{MU} = 89.46%, almost equivalent to Barr and Linebarger's (2008) finding of 91%). Anderson and Levin (1976) and Bickham et al. (2001) both found music to be one of the most highly salient variables in their respective content analyses, eliciting high attention for 24-month-olds; these same two sources also found that attention was highest to rapid pacing, overall. Bryant et al. (1981) and Schmitt et al. (1999) noticed particular orienting to lively music sources over calm ones, direct evidence for the infant fixation on perceptual salience. As a result, high attention to music, especially throughout the entire clip, was expected due to its saturated salience and complexity.

Music had, unsurprisingly, the most pronounced negative effect on learning. A small subset of the literature on features has placed particular emphasis on music, specifically because of its frequent use in commercial sources and its connotation as a highly artistic and intellectual mode of expression. Most of these sources have found the attraction elicited by music to be detrimental to understanding content, which creates attentional competition between the two (Barr & Linebarger, 2008). Barr et al. (2010b) and Barr and Linebarger (2008) both overlaid a music track on the same imitation paradigm with puppets and found similar effects: compared to a non-music video condition, infants performed worse on the deferred imitation tasks than when the music track was imposed; both values were significantly below music- and non-music live demonstrations. The presence of music in the live condition was slightly below the live baseline, but not significantly so (Barr & Linebarger, 2008; Barr et al., 2010b). As a result, music was judged to worsen the video deficit effect, and the current research supports this assertion.

Because background music appears to have little effect in live demonstrations, it presumably plays a role in how information is processed uniquely from a two-dimensional source. DeLoache's (1987) dual representation hypothesis argues that symbolic transfer is more difficult when the attention to the stimulus distracts from this already difficult task. Indirectly, therefore, since the salience of music makes it difficult to inhibit (falling outside of the infant's developmental capacity), representational transfer is impaired. In essence, the high salience and continuous presence of music throughout the entire clip distracts the infant from learning from the television, despite his or her attention to it. Barr et al., (2010b) suggests that music adds significant cognitive load to the infant brain during the demonstration, and as such, the infant must centrate on variable (Piaget, 1930). As reflected by current results, music will be selected over the instruction because of its higher salience quotient.

Another theoretical reason the Music trial was not facilitative is because since it is often present throughout the entire presentation as a background track (54.05% of the time, from the content analysis), the feature sampling model (Huston & Wright, 1983) is presumably inapplicable. Barr et al. (2009) emphasized purposefully matching features to important content in presentation to mark important information. Because uninterrupted music, unlike discrete sound effects, was played throughout the entire clip, music could not be used as a reliable cue to help encode specific target actions more thoroughly. Thus, the Music condition, as one of the most pervasive formal features, yielded the weakest performance and had the most negative influence on learning from video.

Articulating the Similarities and Differences

Overall, Hypothesis 3 was supported because condition differences existed between numerous condition pairings, enough so that a logical ranking system could be determined based on their specific effects (Table 3). Two key motifs can be gleaned from these results: 1) the specific effect of each condition is not dependent on sensory modality (i.e. auditory versus visual), and 2) attention does not markedly predict comprehension.

Since both auditory conditions (Cartoon Voice and Music) were ranked first and last, respectively, with the two visual conditions (Visual Effects and Artificial Dynamics) sandwiched in the middle, it is clear that sensory modality is not a mediating variable in formal feature influence. The way in which information enters the child's brain is not stronger through the eyes over the ears, or vice versa. From these findings, it seems that presentational aspects such as perceptual salience (visual and auditory), pace, continuity, consistency (matching features to content) and intent (are they purposefully placed to enhance content, or not?) are what improves a formal feature's likelihood of being helpful rather than harmful.

Thus, except for the Cartoon Voice condition that maintained learning ability, formal features were detrimental to the currently evidenced ability to learn from video. Adding formal features introduced a layer of sensory complexity that viewers can only supersede if it fits within their intellectual capacity ('age appropriate'). Arguably, if a formal feature displaces learning content outside of this capacity, it will be detrimental, and if it pushes it further in, it will be facilitative (Valkenburg & Vroone, 2004).

Still, the larger question is how formal features are able to do this, in either direction. It has already been determined that, despite their variation, formal features are conventions of an artificial medium and not generally present in real life. What is clear from the data is that the attention directed towards formal features is not a reliable predictor of their effect on comprehension. Results from this study indicated that formal features do not enhance attention to the screen; rather, they maintain the already high fixation to television as a whole. Thus, Hypothesis 1 was not supported, but Hypotheses 2 and 3 were— namely, infants had universally high levels of attention, but varying levels of performance, with formal feature conditions (mostly) performing worse than controls. Therefore, it is clear that while many different kinds of stimuli may incite an orienting response, in some conditions, attention keeps the viewer focused, and in others, the same source of that attention distracts (Anderson et al., 1981). It is important to note that infants' proficiency with the task cannot be reduced to the fact that they weren't given the chance to encode information. As Table 1, indicated, across all conditions, over 2/3 of participants watched both critical actions over every condition, and only 2 participants had only one trial where they witnessed neither. Infants had ample time to observe the target action; lack of exposure to the demonstration was not what affected performance.

Thus, as this analysis has explained, the video deficit effect, (Macklin, 1994; Murray & Murray, 2008; Barr et al. 2009; Anderson & Hanson, 2010; Courage & Setliff, 2010) the feature sampling model (Huston & Wright, 1983; Courage and Setliff, 2010), perceptual impoverishment theory (Barr et al., 2007), dual representation (DeLoache, 1987), the active and passive models (Anderson et al., 1981), and the moderate discrepancy hypothesis (Valkenburg & Vroone, 2004) are all implicated in how formal features can affect performance, if they are applied appropriately. Across the board, the effectiveness of formal features appears to be dependent on how well they can adapt themselves to the viewer's rapidly developing knowledge and experience. Therefore, the crux of formal features is not just what they are, but how they are used.

Practical Implications

Practically, these conclusions have vast implications for understanding how media influence children. Since production companies have been so reprimanded for their ruthless targeting of 'helpless' viewers, research such as this can help clarify to what extent, and in which ways instructional media aimed at infants actually affects learning. As the market for infant learning media grows, so does the voice of their combatants; the polemical movement against The *Baby Einstein* Company from Campaign for a Commercial-Free Childhood (CCFC) has purported an ongoing legal and social battle between two parties who believe their efforts are the most beneficial for healthy development. On one side, media companies are criticized for false advertising of the educational and cognitive benefits of their products, but on other, advocacy groups have just as much trouble finding reputable information to substantiate that media is as detrimental as their cause espouses. Inaccessibility and misattribution of research has hindered these conflicts from stabilizing. Therefore, both empirical researches with televised stimuli (such as the current study), as well as research with commercial material need to continue in order to better substantiate these arguments.

While these results provided ample insights on how the video influences learning, there were many limitations to the current study. As a survey-style analysis, only very specific formal features were adapted to the paradigm. While standardizing these variables helped isolate unique condition effects, these findings only superficially represented their respective family of features. Each condition in this study (Cartoon Voice, Visual Effects, Artificial Dynamics, and Music) has myriad possibilities for how they can be manifested in children's media, and each merit their own comprehensive study to further elucidate their unique condition effects on development. Given the constraints of the researchers' ability and the programs used to make the stimuli, more sophisticated tools would have allowed for the imposition of formal features to appear more like they do in commercial content. As with many studies of this nature, highly controlled lab conditions often run the risk of reducing reliability. Another limitation resulted from the repeated measures design— while advantageous in that variance was reduced in showing all conditions to all participants, participants ran the risk of learning the task from repeating it, though overall, order effects were not observed. Lastly, small oversights from the process, such as removing commercial presence from stimuli (some mothers noted that their child recognized the Cookie Monster doll, perhaps biasing them), should be noted in the future.

As technology and media outlets become more sophisticated and diverse, the scope of testable variables will increase. A field that was once mainly preoccupied with television has now expanded to include computers, video games, new media, electronic toys, and even virtual reality and 3D presentations (Strasburger et al., 2009). While children of different age groups may naturally encounter some of these media outlets more than others, only by studying them all will media's highly saturated presence be understood more fully. A recent study by Zack, Barr, Gerhardstein, Dickerson, and Meltzoff (2009) observed imitation from novel touch-screen technology, revealing new insights on the difficulty between two dimensional transfer, even when children are allowed more interactivity with the screen. Studies such as Zack et al. (2009) reflect the rapidly changing media landscape and the child's place within it.

Given this knowledge, the current research has many potential future directions. As mentioned, the diversity and range of formal features present in children's media was only barely probed in this study. As reflected in previous literature, each category and type of formal features deserves further attention and specified work (for example, a study observing the role of interspersed versus continuous background music). Still, even with the current data, there are many possibilities for further analyses. For example, given the survey demography, a potential future study could compare results between two groups based on daily television consumption (using a median split) into two operationalized groups: 'frequent watchers' and 'low watchers.' Completing this study could shed more light on the role that relative experience with media conventions plays in processing of audiovisual stimuli. On a larger scale, replicating this work with a between-subjects design would also allow for more statistical tests to enhance validity.

The role of media in the youngest, and arguably most mollifiable, child's life still incites many questions that have yet to be answered. Nonetheless, the current work on formal features has helped unify the vital roles that appearance, content, and intent all play in communicating information to the viewer. As these opportunities for insight grow with society, technology, and research, a more solid understanding of how the child and the screen interact will eventually be achieved.

Acknowledgements

The design, creation, and implementation of this research would not have been possible without the generous support of Dr. Emily W. Bushnell, my academic advisor and professor of Psychology at Tufts University. Her willingness to help with all aspects from the study, from discussing relevant literature, to siting on the ground and helping me build stimuli, has been pivotal to the success of this project. Also very influential was the guidance and advice of Dr. Julie Dobrow, head of the Communications and Media Studies department at Tufts University. Her perspective as an educator and advocate of media literacy helped inform my work from a larger perspective outside the laboratory.

Thank you also to Bianca Velayo and Kathleen Holec, my friends and associates in the Behavioral Development Lab. Their constant attention and help on all aspects of this study was vital to it reaching its full potential. Finally, I would like to thank my family and friends for their constant support throughout the largest endeavor of my academic career to date. Without their words of encouragement, this project never would have been what it became.

This research was funded by the Undergraduate Research Fund at Tufts University; their support provided the necessary materials to bring all of the design elements of this project to life.

References

American Academy of Pediatrics. (1999). Media education. *Pediatrics*, *104*(2), 341-343.

Anderson, D. R. & Bryant, J (1983). Research on television viewing: The state of the art. In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of TV: Research on attention and comprehension*. New York, NY: Academic Press.

Anderson, D.R. & Hanson, K.G. (2010). From blooming, buzzing confusion to media literacy: Early development of television viewing. *Developmental Review, 30,* 239-255.

Anderson, D. R., & Kirkorian, H. L. (2006). Attention and television. In J. Bryant, & P. Vorderer (Eds.), *The psychology of entertainment* (35–54). Mahwah, NJ: Erlbaum.

Anderson, D. R., & Levin, S. R. (1976). Young children's attention to Sesame Street. *Child Development*, *47*, 806-811.

Anderson, D. R., & Lorch, E. P. (1983). Looking at television: Action or reaction. In J. Bryant & D. Anderson (eds.), *Children's understanding of television* (1-33). New York, NY: Academic Press.

Anderson, D. R., Levin, S. R., and Lorch, E. P. (1977). The effects of television program pacing on the behavior of preschool children. *AV Communication Review, 25*, 159-166.

Anderson, D. R., Alwitt, L. F., Lorch, E. P., and Levin, S. R. (1979). Watching children watch television. In G. Hale and M. Lewis (Eds.), *Attention and Cognitive Development*. New York: Plenum.

Anderson, D. R., Lorch, E. P., Field, D. E., & Sanders, J. (1981). The effect of television program comprehensibility on preschool children's visual attention to television. *Child Development*, *52*, 151–157.

Barr, R., & Hayne, H. (1996). The effect of event structure on imitation in infancy: Practice makes perfect? *Infant Behavior and Development*, 19, 255-259.

Barr, R., & Hayne, H. (1999). Developmental changes in imitation from television during infancy. *Child Development*, 70, 1067-1081.

Barr, R. & Linebarger, D.L. (2008). Music interferes with learning from television during infancy. Philadelphia, PA: Annenberg School for Communication, University of Pennsylvania.

Barr, R., & Wyss, N. (2008). Reenactment of televised content by 2-year-olds: Toddlers use language learned from television to solve a difficult imitation problem. *Infant Behavior and Development*, 31, 696–703.

Barr, R., Muentener, P., Garcia, A., Fujimoto, M. & Chavez, V. (2007). The effect of repetition on imitation from television during infancy. *Developmental Psychobiology*, *49(3)*, 196-207.

Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008). Infants' attention and responsiveness to television increases with prior exposure and parental interaction. *Infancy*, *13*, 30–56.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development*, *19*, 159-171.

Barr, R., Somanader, M., & Wyss, N. (2009). Imitation from television during infancy: The role of sound effects. *Journal of Experimental Child Psychology*, *123*, 1–16.

Barr, R., Lauricella, A., & Zack, E. (2010a). Infant and early childhood exposure to adult-directed and child-directed television programming relations with cognitive skills at age four. *Merrill-Palmer Quarterly*, *56(1)*, 21-48.

Barr, R., Shuck, L., Salerno, K., Atkinson, E., & Linebarger, D. L. (2010b). Music interferes with learning from television during infancy. *Infant and Child Development*, *19*, 313-331.

Beentjes, J. W. J., de Koning, E., & Huysmans, F. (2001). Children's comprehension of visual formal features in television programs. *Journal of Applied Developmental Psychology*, *22*, 623–638.

Bickham, D. S., Wright, J. C., & Huston, A. C. (2001). Attention, comprehension, and educational influences of television. In the D. G. Singer & J. L. Singer (Eds.), *Handbook of children and the media* (pp. 101–119). Thousand Oaks, CA: Sage.

Bryant, J., Brown, D., Silverberg, A. R. & Elliott, S. M. (1981). Effects of humor in advertising. *Human Communication Research*, *8(1)*, 43-57.

Christakis, D.A., Zimmermann, F.J., DiGiuseppe, D.L., & McCarty, C.A. (2004). Early television exposure and subsequent attentional problems in children. *Pediatrics*, *113(4)*, 708-713.

Courage, M. L. & Setliff, A. E. (2010). When the television is on: the impact of infant directed video on 6- and 18-month-olds' attention during toy play and on parent infant interaction. *Infant Behavioral Development*, *33(2)*, 176-88.

Courage, M., Murphy, A. N., Goulding, S., & Setliff, A. E. (2010). When the television is on: The impact of infant-directed video on 6- and 18-month olds' attention during play and on parent-infant interaction. *Infant Behavior and Development*, *33*, 176-188.

DeLoache, J. S. (1987). Rapid chance in the symbolic functioning of very young children. *Science*, *238*, 1556-1557.

Flavell, J. H., Flavell, E. R., Green, F. L., & Korfmacher, J. E. (1990). Do young children think of television images as pictures or real objects? *Journal of Broadcasting and Electronic Media*, *34*, 399–419.

Flynn, E., & Whiten, A. (2008). Imitation of hierarchical structure versus component details of complex actions by 3- and 5-year-olds. *Journal of Experimental Child Psychology*, 101, 228–240.

Greer, D., Potts, R., Wright, J. C., & Huston, A. C. (1982). The effects of television commercial form and commercial placement on children's social behavior and attention. *Child Development,* 53, 611-619.

Gunter, B. & McAleer, J. (1990) *Children and Television (2nd Ed.)*. London, England: Routledge.

Hayne, H., Barr, R., & Herbert, J. (2003a). The effect of prior practice on memory reactivation and generalization. *Child Development*, 74, 1615–1627.

Hoffner, C., & Cantor, J. (1985). Developmental differences in responses to a television character's appearance and behavior. *Developmental Psychology*, *21*, 1065-1074.

Huston, A. C., & Wright, J. C. (1983). Children's processing of television: The informative function of formal features. In Bryant, J. & Anderson, D. R. (Eds.) *Children's Understanding of Television*. New York, NY: Academic Press, 35-68.

Lemish, D. (2007) *Children and television: A global perspective.* Oxford UK: Blackwell Publishing.

Macklin, M. C. (1994). The effects of an advertising retrieval cue on young children's memory and brand evaluations. *Psychology and Marketing*. *11(3)*, 291-311.

Macklin, M. C. (1996). Preschoolers' learning of brand names from visual cues. *Journal of Consumer Research*, *23(3)*, 251-261.

McCall, R. B., Parke, R. D., & Kavanaugh, R. D. (1977). Imitation of live and televised models by children one to three years of age. *Monographs of the Society for Research in Child Development, 42*(5), 1-94.

Meltzoff, A. N. (1988). Imitation of televised models by infants. *Child Development*, *59*, 1221-1229.

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario," *Child Development*, *74 (1)*, 221-237.

Murray, J. P., & Murray, A. D. (2008). Television: Uses and effects. In M. M. Haith & J. B. Benson (Eds.). *Encyclopedia of infant and early childhood development*, *3*, 309-318.

Neeley, S. M. & Schumann, D. W. (2004). Using animated spokes-characters in advertising to young children: Does increasing attention to advertising necessarily lead to product preference? *Journal of Advertising*, *33(3)*, 7-23.

Piaget, J. (1930). *The child's understanding of physical causality*. Oxford, England: Harcourt, Brace.

Pierroutsakos, S. L. & Troseth, G. L. (2003). Video verité: Infants' manual investigation of objects on video. *Infant Behavior and Development*, *26*, 183-199.

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: behavioral and psychophysiological measures. *Developmental Review*, *30*, 203-219.

Richards, J.E. & Anderson, D.R. (2004). Attentional inertia in children's extended looking at television. In R.V. Kail (Ed.), *Advances in Child Development and Behavior* (Vol. 32 pp. 163-212), Amsterdam: Academic Press.

Rideout, V., & Hamel, E. (2006). *The media family: Electronic media in the lives of infants, toddlers, preschoolers and their parents.* Menlo Park, CA: Kaiser Family Foundation.

Robb, M. B., Richert, R. A., & Wartella, E. A. (2009). Just a talking book? Word learning from watching baby videos. *Developmental Psychology*, *27(1)*, 27-45.

Roedder, D. L. (1981). Age differences in children's responses to television advertising: An information processing approach,". *Journal of Consumer Research*, *8*, 144-153.

Schmitt, K.L., Anderson, D.R. & Collins, P.A. (1999). Form and content: Looking at visual features of television. *Developmental Psychology*, *35,* 1156-1167.

Schmitt, K.L., Anderson, D.R., & Collins, P.A. (1999). Form and content: Looking at visual features of television. *Developmental Psychology, 35*, 1156-1167.

Strasburger, V. C., Wilson, B. J., & Jordan, A. B. (2009). *Children, Adolescents, and the Media*. Thousand Oaks, CA: Sage.

Seehagen, S. & Herbert, J. S. (2010). The role of demonstrator familiarity and language cues on infant imitation from television. *Infant Behavior and Development*, *33*, 168-175.

Suddendorf, T. (2003). Early representational insight: Twenty-four-month-olds can use a photo to find an object in the world. *Child Development*, *74*, 896–904.

Strouse, G. A. & Troseth, G. L. (2008). "Don't try this at home": Toddlers' imitation of new skills from video. *Journal of Experimental Child Psychology*, *101*, 262-280.

Susman, E. J. (1978). Visual and verbal attributes of television and selective attention in preschool children. *Developmental Psychology*, *14(5)*, 565-566.

Troseth, G. L. (2003a). Getting a clear picture: young children's understanding of televised image. *Developmental Science*, *6(3)*, 247-253.

Troseth, G. L. (2003b). TV guide: Two-year-old children learn to use video as a source of information. *Developmental Psychology*, *39*, 140–150.

Troseth, G. L., & DeLoache, J. S. (1998). The medium can obscure the message: Young children's understanding of video. *Child Development*, *69*, 950–965.

Troseth, G. L., Saylor, M. M., & Archer, A. H. (2006). Young children's use of video as a source of socially relevant information. *Child Development*, *77*, 786–799.

Turnure, J. E. (1970). Children's reactions to distractors in a learning situation. *Developmental Psychology*, *2*, 115–122.

Valkenburg, P. M. and Cantor, J. (2001). The development of a child into a consumer, in *Children in the Digital Age: Influences of Electronic Media on Development*, S. L. Calvert, A. B. Jordan, and R. R. Cocking (Eds.) Westport, CT:. Praeger, 201-214.

Valkenburg, P. M., & Vroone, M. (2004). Developmental changes in infants' and toddlers' attention to television entertainment. *Communication Research*, *31(3)*, 288–311.

Wartella, E. A. & Ettema, J. S. (1974). A cognitive developmental study of children's attention to television commercials. *Communication Research*, *5*, 69-88.

Wright, J. C., Huston, A. C., Ross, R. P., Calvert, S. L., Rolandelli, D. I., Weeks, L. A., Raeissi, I., & Potts, R. (1984) Pace and continuity of television programs: effects on children's attention and comprehension. *Developmental Psychology, 20*, 653-666.

Zimmerman, F. J., Christakis, D. A,, & Meltzoff, A. N. (2007). Television and DVD/video viewing by children 2 years and younger. *Archives of Pediatric and Adolescent Medicine, 161,* 437-479.

***Appendix A*. Preliminary Feature Analysis**

***A.1.* Sample**

- *Baby Einstein: Visual and Multilingual Experience* (1998)
- *Baby Galileo: Discovering the Sky* (2003)
- *Baby Mozart: Music Festival* (1999)
- *Brainy Baby: Right Brain* (1999)
- *Sesame Street: Elmo's World, Dancing, Music, and Books* (2001)
- *Sesame Street: Play with Me Sesame (2007)*
- *Baby See 'n Sign: Volume I (2004)*
- *Teaching Signs for Baby Minds: Concepts and Combinations (2006)*
- *Teaching Signs for Baby Minds: Everyday Signs (2006)*

***A.2.* Coding Rubric**

I. Voice Over

1. Presence of a Voice Over (narration)
   a. Present
   b. Not present

2. Age and Gender of Voice-Over
   a. Adult male's voice on voice-over
   b. Adult female's voice on voice-over
   c. Child male's voice on voice-over
   d. Child female's voice on voice-over
   e. Non-Human voice-over
      i. Cartoon voice
      ii. Mechanical voice
      iii. Foreign language voice
      iv. Other voice

Tone of Voice-Over
   a. Gentle
   b. Exciting/intense
   c. Happy
   d. Goofy/silly
   e. Somber
   f. Scary
   g. Angry/tough
   h. Other

II. Music

1. Music present in clip
    a. Present
    b. Not present

2. Amount of music
    a. In seconds
    b. Percentage of entire clip

3. Type of music
    a. Vocal, sung by kids
    b. Vocal, sung by adults
    c. Vocal, sung by both kids and adults
    d. Instrumental
    e. Both vocal and instrumental

4. Distribution of music
    a. Throughout
    b. Interspersed

5. Speed of music
    a. Faster than normal
    b. About normal
    c. Slower than normal

6. Volume of music
    a. Louder than normal
    b. About normal
    c. Softer than normal

7. Tone of music
    a. Gentle
    b. Exciting/intense
    c. Happy
    d. Goofy/silly
    e. Somber
    f. Scary
    g. Angry/tough
    h. Other

III. Sound Effects (non-Musical)

1. Sound effects used
    a. Throughout
    b. Interspersed
    c. Not at all

2. Types of Sounds used
    a. Explosions (bang, boom, etc.)
    b. Chimes/bells/jingling
    c. Bounce/springs/hops
    d. Horns/honks
    e. Whistles
    f. Popping
    g. Wheezing/crackling
    h. Exclamations (wow! wee! etc.)
    i. Other

3. Tone of sounds used
    a. Gentle
    b. Exciting/intense
    c. Happy
    d. Goofy/silly
    e. Somber
    f. Scary
    g. Angry/Tough
    h. Other

4. Speed of sound effects
    a. Faster than normal
    b. About normal
    c. Slower than normal

5. Volume of sound effects
    a. Louder than normal
    b. About normal
    c. Softer than normal

IV. Visual Effects

1. Visual effects used
    a. Throughout
    b. Interspersed
    c. Not at all

2. Cartoon use
    a. No cartoons
    b. Some cartoons used (interspersed)
    c. Some cartoons used (integrated)
    d. All cartoon

3. Puppet use
    a. No puppets
    b. Puppet narrator
    c. Puppet supporting characters
    d. Puppets used only once

4. Types of visual effects (mark all that apply)
    a. Distortions
    b. Flying objects
    c. Flashes
    d. Shapes on screen
    e. Object color changes
    f. Letters/words on screen

5. Color salience
    a. Black and white
    b. Muted coloring
    c. About normal coloring
    d. Intense, bright coloring
    e. Changes in overall coloring throughout

V. Pacing

1. Rapidity of Action
    a. Faster than normal
    b. About normal
    c. Slower than normal

2. Volume of Action
    a. Louder than normal
    b. About normal
    c. Softer than normal

3. Description of Movement
    a. Normal
    b. Energized
    c. Jerky/erratic
    d. Fluid
    f. Other

IV. Camera Changes

1. Zoom Effects
    a. No zoom effects (stable camera)
    b. Some zooming
    c. Frequent zooming

2. Camera Angles
    a. No angle changes (stable camera)
    b. Some slow angle changes
    c. Some fast angle changes
    d. Some of both
    e. Frequent slow angle changes
    f. Frequent fast angle changes
    g. Both Frequently

A.3. **Variable Operationalization**

- In general, 'frequent' refers to an occurrence of roughly more than 50% of the clip, sometimes refers to an occurrence of roughly 25-50% throughout the clip, and 'never' refers to 0% occurrence

I. Voice-Over

- A **voice-over** shall be defined as narrative presence, which, to a variable degree, explains the purpose of what is being depicted on screen. The source of the voice, while its identity may be known or recognizable, must be disembodied and its source must not be shown in any capacity. The voice over may be a constant narration, or may be interspersed throughout the clip.

- **2.e.i. Cartoon voice**: this categorization refers to a voice that while, produced by a human, is reminiscent of classic animated characterizations, specifically exemplified by, but not limited to, by any or all of the following: inhumanly unusual pitch (high or low), unnatural pacing (too fast or too slow), jaunty, almost musical timbre, nonsensical or outlandish quips, and a generally amusing and silly tone; portrayals may range from the giggly and squeaky tone of an anthropomorphic mouse or fairy, to the gruff babblings of a friendly monster; must be saccharine and artificial in demeanor; any stilted, mechanical, robotic or non-organic sounding voices are excluded

- **2.e.ii. Mechanical voice**: this characterization refers to any non-human, artificial sounding voice that does not fall into the cartoon voice category, primarily encapsulating that which is presumed to be technologically inspired, including, but not limited to: stilted pauses between words or sentences, fragmented prosody, computerized/digitized speech, and mispronunciations often characteristic of both real and fantastical robots and mechanical beings

- **4.** The following delineates criteria for what features constitute a particular tone, when used in a spoken or musical context:

  a. **Gentle**: the tone of voice or music is generally soft, pleasant, and is meant to elicit a soothing or calming effect on the viewer; the timbre and prosody flow lightly and easily

  b. **Friendly/Happy**: the tone of voice or music is pleasant, but more upbeat and jovial than the gentle category; it may range from warm and inviting to joyful and exuberant, without neither reaching the tender quietude of the gentle category, or the abnormal excitement of the exciting/intense category

  c. **Goofy/silly**: the tone of voice or music is characterized by the implication of exaggerated jest or ridicule, with the intention of making the viewer laugh; qualities include, but are not limited to: abnormally distorted voices to a comical effect, bouncy, jaunty tunes, and exaggerated onomatopoeic references meant to mimic or imitate others in jest (e.g. chuckling laughter, flatulence, endearing sarcasm, nonsense words)

  d. **Exciting/intense**: the tone of voice or music is characterized by either fast paced, loud exhilaration and enthusiasm, or both, ranging to the level of hysteria; vocal cues may be overwhelming at times, and must maintain a consistent level of high amplification; fundamentally designed to arouse and enhance the attention of the viewer

  e. **Somber**: the tone of voice or music is directly antithetical to the exciting/intense category, with features exemplified by either slow paced, low volume, or both; the intent of the depiction is clearly to make the viewer melancholy, sad, or contemplative

  f. **Scary**: the tone of voice or music is specifically characterized by forms that are clearly meant to frighten or chill the viewer; particular instances include, but are not limited to: exaggerated foreign accents and 'monstrous' voices (e.g. a la Dracula, Frankenstein), hooting, howling, screaming, eerie silences, and sudden changes in pace or volume with the intent to startle; excessive yelling and gruffness should be considered under the angry/tough category, not the scary category

  g. **Angry/tough**: the tone of voice or music is reflects aggression and violence, in such manifestations as loud voice, rude behavior, pejorative attitude and brutish demeanor, but may be as diminutive as simply projecting a competitive streak or confident cockiness with the speaker's prosody and timbre; vocal and musical cues may be fast and loud, but will fall into the exciting/intense category unless they suggest aggression of any kind

h. **Other**: a tone of voice or musical quality that falls radically out of any of the aforementioned categorizations

Pacing Changes

- **4.** Description of movement: this category is meant to be an overall qualifier for the way in which the pacing of the video stimulus affects the movement quality of the action

  a. **Normal** refers to a typical movement quality that seems unaffected by any unusual effects, and does not fit into any of the following categories

  b. **Energized** refers to movement that suggests excitement and exhilaration, usually categorized by rapid pacing, including, but not limited to: deliberate fast forwarding

  c. **Jerky/erratic** refers to movement that is specifically categorized by inconsistent, non-fluid motion and jumps from one action to the next, almost, including, but not limited to: pauses, rapid changes from fast to slow, interspersed fast forwarding, slow motion, or both; action may appear energized, but only falls into this category if the movement is inconsistent

  d. **Fluid/calm** refers is directly antithetical to the jerky/erratic category, with movement characterized by lyrical and flowing movements with absolutely no stops, pauses, or changes in pacing; usually slow and methodical, but doesn't have to be, with techniques including, but not limited to: deliberate slow motion

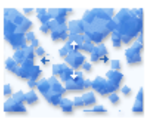
  e. **Other** refers to a movement quality that falls radically out of any of the aforementioned categorizations

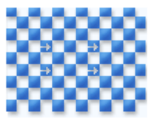
Appendix B. Condition Schemata

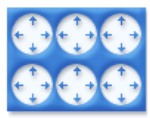
B.1. Cartoon Voice Transcript

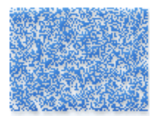
"Hi baby, how are you doing? Look at me! Wow, look at all these toys, look! Wow look at that! All right baby watch me again; look at this! Wow, gee-whiz! All right baby look at me, one more time cone one! Wow look at that! All right baby, we're going to do it again, watch me, look! Wow, gee-whiz look at that! All right baby look, one more time look! Wow, look at that! All right baby, last time, watch me, look t this! All right baby last time, look at this! Wow, gee-while! All right baby, see you later, thanks, bye!"

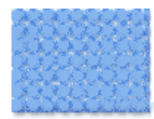
B.2. Visual Effect Transitions (in order*)*


Shatter, In


Checkerboard, Across


Circles


Dissolve


Whirlwind


Stars, 5 Points

B.3. Artificial Dynamics Segments

WHERE TRIALS		
Segment	Original	Final
Slow 1	5	10
Freeze 1	0	2
Slow 2	5	10
Freeze 1	0	2
Fast 1	7	1.75
Freeze 3	0	2
Fast 2	7	1.75
Freeze 4	0	2
Slow 3	5	19
Freeze 5	0	2
Fast 3	18	4.5
Slow 4	1	1
TOTAL	≈ 51 sec	≈ 58 sec

WHICH TRIALS		
Segment	Original	Final
Slow 1	7	10
Freeze 1	0	2
Slow 2	7	10
Freeze 1	0	2
Fast 1	8	1.75
Freeze 3	0	2
Fast 2	20	1.75
Freeze 4	0	2
Slow 3	7	19
Fast 3	2	2
	18	4.5
	1	1
TOTAL	≈ 48 sec	≈ 50 sec

The left columns indicate the modified effect (Slow = 0.5x; Fast = 4x; Freeze = freeze frame), the middle columns denote after how many seconds the clip was parsed, and the right columns show the resulting length of each segment after modification. Separate, but comparable rubrics were done for WHERE vs. WHICH to give equal treatment of slow and fast motion to each object as well as both critical actions .

Appendix C. Parent Survey

PARENT SURVEY

ease answer the following to the best of your ability. If you have any questions or concerns, feel free to ask a researcl assistant after the study has been completed. Please be assured that your answers will not be linked to you or your child's identity, and will only be used for statistical purposes. Thank you very much for your input!

ART I: Please fill in the blank next to each question; if a question does not apply, please write "N/A":

1. How old is your child (in months)? __________
2. Does your child have any older siblings? If so, how old are they? __________
3. On average, how many hours per day is your child exposed to television/video? __________
4. Have you ever used early childhood learning videos (e.g. *Baby Einstein*, *Brainy Baby*, *Baby Sign*)? ______
   a. If so, which one(s)? ____________________________________
   b. If not, why not? ____________________________________
5. Do you subscribe to any early childhood learning channels such as *BabyFirst TV* or *Sprout*? __________
   a. If so, which one(s)? ____________________________________
   b. If not, why not? ____________________________________

ART II: Please circle the number that best describes your agreement with the following statements:

1. I believe that children shouldn't be exposed to television/video before a certain age

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

2. I think that early learning from television/video can be beneficial to my child's overall development

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

3. I think that television/video for children before elementary school age should be only educational

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

4. I have been or would be interested in regularly viewing early childhood learning channels such as *BabyFirst TV* and *Sprout* with my child if they were readily accessible

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

5. I have been deterred from looking into early childhood video companies (e.g. *Baby Einstein*, *Brainy Baby*, *Baby Sign*) because they are surrounded by controversial findings and negative publicity

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

6. I think advertising to children is disruptive and harmful to individual development

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

7. Enforcing specific television/video watching rules in our household is a priority

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

8. I believe in the importance of viewing television with my child, so as to promote early media literacy (i.e. the developing ability to be cognizant of the intentions, content, and function of what one is viewing)

1	2	3	4	5	6	7
Strongly disagree			**Neutral**			**Strongly agre**

Appendix D. Data Scoring Forms

A

SCORING (POST)
Date of Scoring:

TELEVISED FORMAL FEA
Tufts University Behavioral Develo
Marco
Honors Thesis 2

PARTICIPANT NO.:
Date:
Date of Birth:
Age (months/days):
Gender:

EXPERIMENTERS:
Primary Tester:
Assistant:

I. Looking Time (VIDEO only, in seconds)

TRIAL	CONDITION	TIME	CLIP	% ATT	LOOKED CA?	
Warm-Up	—		—	—	**1**	**2**
T1						
T2						
T3						
T4						
T5						
T6						

* CA = **critical action** (WHERE: 1 = placement of hidden toy, 2 = second showing of hidden toy)
(WHICH: 1 = first showing of chosen toy, 2 = second showing of chosen toy)

II. Qualitative Observations (VIDEO and TOYS**)**

TRIAL	CONDITION	NOTES
Warm-Up	—	
T1		
T2		
T3		
T4		
T5		
T6		

Additional Notes:

USE	REJECT

PARTICIPANT NO.:

Date:

Date of Birth:

Age (months/days):

Gender:

EXPERIMENTERS:

Primary Tester:

Assistant:

TRIAL	CONDITION	NOTES
Warm-Up	—	
T1		
T2		
T3		
T4		
T5		
T6		

Additional Notes:

PARTICIPANT NO.: __________ D.O.B.: __________ Age (m/d): __________ Gender: __________ Primary Tester: ____________

Assistant: ____________

III. Behavior (TOYS only)

T1: _______________ Chosen Toy: _______________

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

T2: _______________ Chosen Toy: _______________

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

T3: _______________ Chosen Toy: _______________

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

A^L = looks at toy on baby's left
A^T = touches toy on baby's left
A^P = picks up toy on baby's left
A^W = TA with toy on baby's left*

B^L = looks at toy in center
B^T = touches toy in center
B^P = picks up toy in center
B^W = TA with toy in center*

C^L = looks at toy on baby's right
C^T = touches toy on baby's right
C^P = picks up toy on baby's right
C^W = TA with toy on baby's right*

F = finds hidden toy (WHERE only)
F^L = looks at hidden toy (WHERE ONLY)
F^T = touches hidden toy (WHERE ONLY)
F^P = picks up hidden toy (WHERE ONLY)

S = makes toy work (WHICH only)

$T^{X\text{-}Y}$ = acts on two toys together
X = action occurs on the chosen toy
•

n.b.: 'X' and 'Y' represent any toy A, B, or C, with X being the toy in the left hand and Y being the toy in the right hand
*TA indicates completion of the target action, regardless of target toy; note other mentionable types of play qualitatively
as a result, for all intents and purposes, P = W for WHERE trials (but not for WHICH), unless otherwise noted

B

1
OVER →

PARTICIPANT NO.: __________ D.O.B.: __________ Age (m/d): __________ Gender: __________ Primary Tester: ____________

Assistant: ____________

T4: _______________ Chosen Toy: _______________

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

T5: _______________ Chosen Toy: _______________

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

T6: _______________ Chosen Toy: _______________

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

A^L = looks at toy on baby's left
A^T = touches toy on baby's left
A^P = picks up toy on baby's left
A^W = TA with toy on baby's left*

B^L = looks at toy in center
B^T = touches toy in center
B^P = picks up toy in center
B^W = TA with toy in center*

C^L = looks at toy on baby's right
C^T = touches toy on baby's right
C^P = picks up toy on baby's right
C^W = TA with toy on baby's right*

F = finds hidden toy (WHERE only)
F^L = looks at hidden toy (WHERE ONLY)
F^T = touches hidden toy (WHERE ONLY)
F^P = picks up hidden toy (WHERE ONLY)

S = makes toy work (WHICH only)

$T^{X\text{-}Y}$ = acts on two toys together
X = action occurs on the chosen toy
•

n.b.: 'X' and 'Y' represent any toy A, B, or C, with X being the toy in the left hand and Y being the toy in the right hand
*TA indicates completion of the target action, regardless of target toy; note other mentionable types of play qualitatively
as a result, for all intents and purposes, P = W for WHERE trials (but not for WHICH), unless otherwise noted

2

C **SCORING (POST)**
Date of Scoring:

TELEVISED FORMAL FEATURES
Tufts University Behavioral Development Lal
Marcos Sastre II
Honors Thesis 2010 -201

PARTICIPANT NO.:
Date:
Date of Birth:
Age (months/days):
Gender:

EXPERIMENTERS:
Primary Tester:
Assistant:

IV. Latency Time until TA (TOYS only; number of seconds since start of trial – CONTINUOUS)

Trial	**Condition**	**Toy**	**A**				**B**				**C**				S_{LAT}
	Chosen	**Action**	L	T	P	TA	L	T	P	TA	L	T	P	TA	
T1															
T2															
T3															
T4															
T5															
T6															

A = toy on baby's left
B = toy in center
C = toy on baby's right
– = actions are linked
* = occurs simultaneously

L = time until look
T = time until touch *(includes L, otherwise note)*
P = time until pick up *(includes L and T, otherwise note)*
TA = time until target action *(code for all, including non-target)*
X = action does not occur

V. Order of Action (TOYS only; in what sequence each toy was acted upon – ORDINAL)

Trial	**Condition**	**Action**	**L**			**T**			**P**			**TA**			S_O
	Chosen	**Toy**	A	B	C	A	B	C	A	B	C	A	B	C	
T1															
T2															
T3															
T4															
T5															
T6															

A = toy on baby's left
B = toy in center
C = toy on baby's right
* = occurs simultaneously

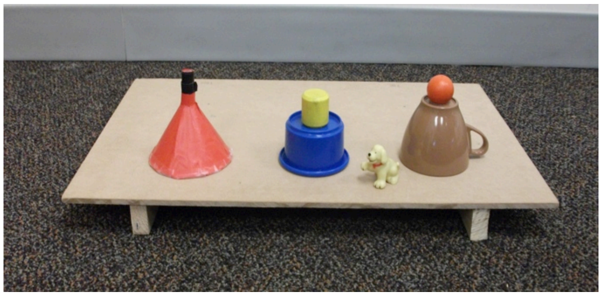
L = looking order
T = touching order
P = picking up order
TA = target action order

1 = first
2 = second
3 = third
4 = does not occur

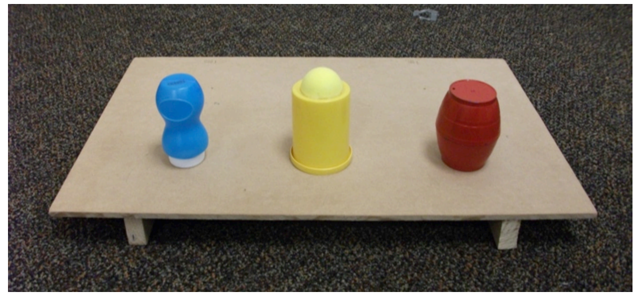
Additional Notes:

Appendix E. Images

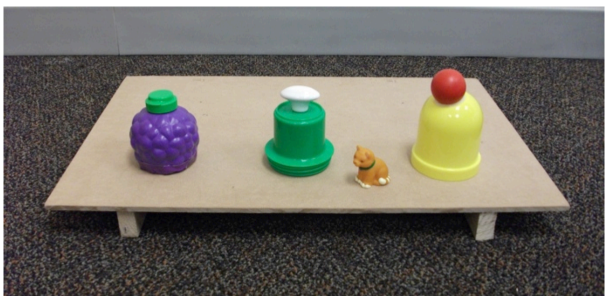
***E.1.* Toy Stimuli**



WHERE 1

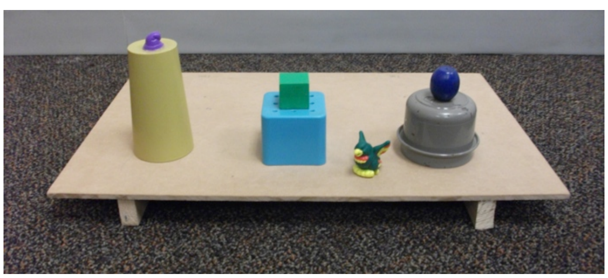


WHICH 4

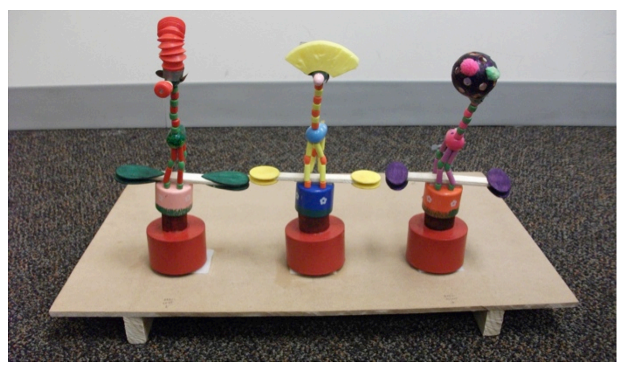


WHERE 2

WHICH 5

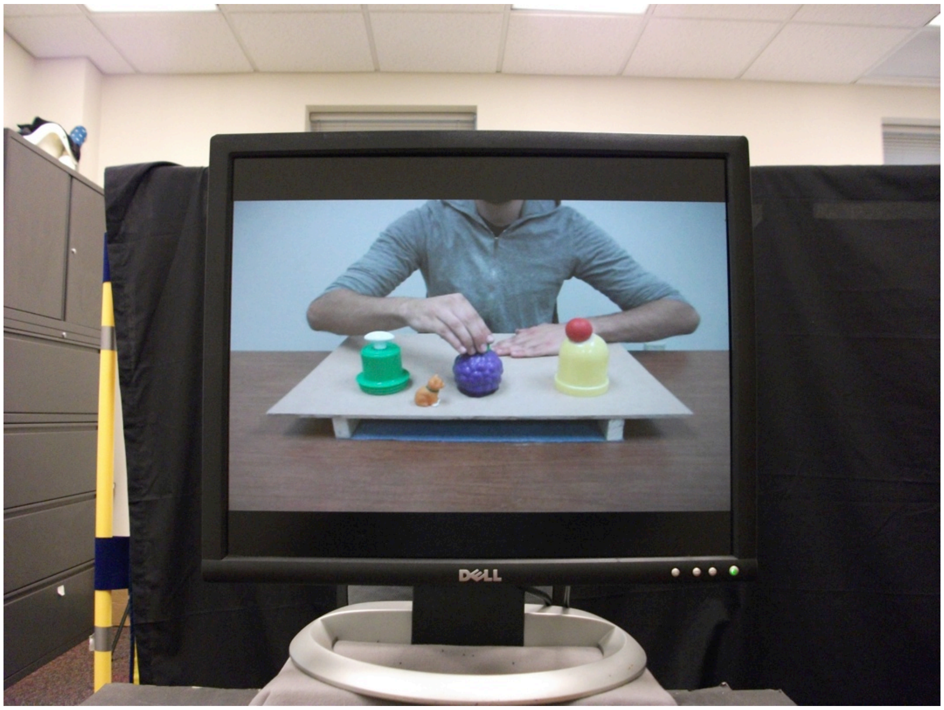


WHERE 3

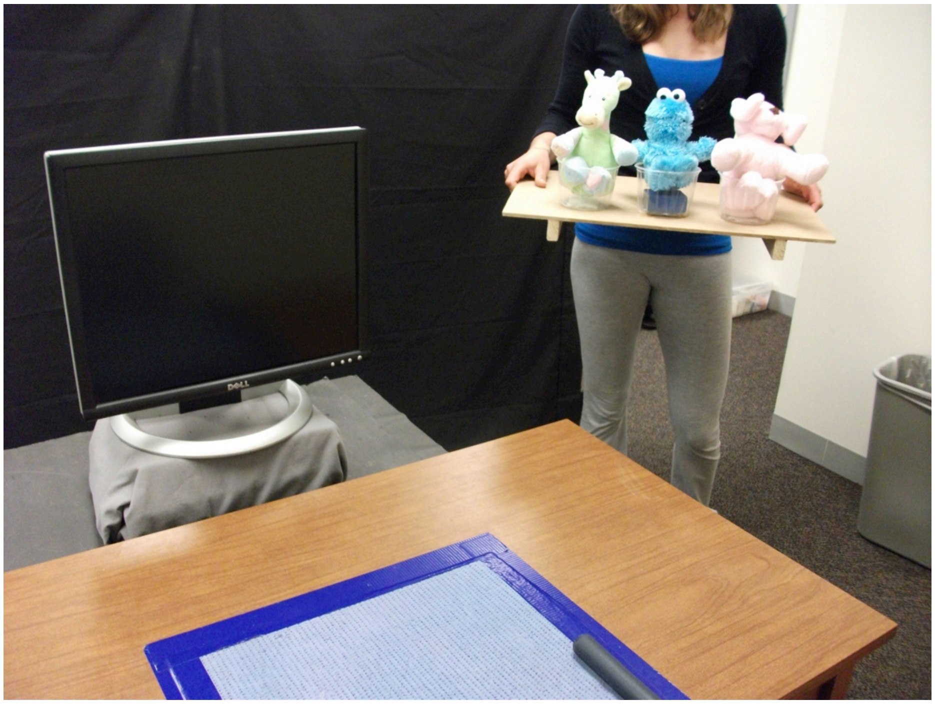


WHICH 6

E.2. **Video Demonstrations**



***E.3.* Presentation of Toys**



***E.4.* Infants Interacting with Video and Toys**

Watching the Video

Completing the Target Action