From World to Word and Back:

An Empirical and Philosophical Investigation on Syntactical Bootstrapping

By

Anjie Cao

Thesis Submitted in Partial Fulfillment of the

Requirements for the

Dietrich College Senior Honors Program

Department of Psychology

Carnegie Mellon University

2020

Acknowledgment

If you told the person I was two years ago that one day she would write a 30,000-word long thesis in English, she would laugh out: “That’s ridiculous!”. But a year ago, when I brought this dual-project idea to my advisors, Dr. David Rakison and Dr. Wayne Wu, neither of them laughed. They sat down with me, brainstormed with me, and guided me through so many confusions and uncertainties. They did so over and over again throughout the summer, the academic year, and the stormy graduate school application season. I am deeply grateful for the opportunities to have learned from and worked with them.

I am thankful for the support I received from the Dietrich College Dean’s Office, especially from Dr. Joseph Devine and Dr. Jennifer Keating. I am honored to receive the Dietrich Honors Fellowship, which enabled me to stay in Pittsburgh and dived into my project during the summer. That immersive experience was invaluable for the development of this project.

I feel so privileged to have studied in the Department of Psychology and the Department of Philosophy at Carnegie Mellon University. I am thankful for Dr. Erik Thiessen, who opened the door to infant research in my freshman summer. I am thankful for Dr. Joel Smith, whose resourcefulness, patience, and kindness lead me into philosophy. I also want to express my appreciation to Dr. Marlene Behrmann, my academic advisor. I look up to her in so many different ways. I can’t list out every professor I have taken classes from, but just as I told my mon in my freshman year: “I feel like going to see live concerts and have face-to-face conversations with rock stars every day!” I still have that excitement with me today as I am approaching my graduation day.

My appreciation also goes to the people I have known through working in Infant Language and Learning Lab and Infant Cognition Lab. I feel incredibly lucky to have known Phoebe Dinh through our Margarete Morrison chat. She is a mentor that has changed my perspectives on many things in very deep and important ways. I would also like to thank my participants and their parents for volunteering their time to support our studies.

Yannan Li, Xiaoxuan Ling, Yiyao Yang – my buddies from 巴沟远征队: well, thanks. As always. We need to, and we certainly will, meet again when this COVID-19 is over, either in Beijing or in the U.S. There are so many things we need to celebrate. I also want to thank my readers in my writing account and Neureality, who stuck with me and my babbling during my senior year. Also, I appreciate the fact that Mitski, Sparklehorse, Thom Yorke, Radiohead, Yo La Tengo, Vampire Weekend, Bon Iver and Lucy Dacus exist in this world. Their music sustained me through difficult times.

Finally, thank you, mom. Thank you for always believing that I deserve the best education, and for always trusting me for making my own decisions. Even during the time everyone else, including me, had doubts.

Part I. Empirical Study

Syntactic Bootstrapping as Noun Learning Mechanism in 12-Month-Olds and 20-Month-Olds Infants

Abstract

Causality is involved in every fabric of our life, but how do young children acquire the concept of causality? Previous studies have shown that infants as young as 6½ months of age can perceive the causal relationship in a simple Michotte launching event. Yet it is not until the age of 2 year that children show evidence of reasoning with causality (Gopnik & Sobel, 2000; Gopnik, Sobel, Shulz & Glymour, 2001; Nazzi & Gopnik, 2003; Sobel, Tenenbaum & Gopnik, 2004). One reason for this lag is language. A conceptual understanding of causality requires understanding causal language, and the developmental trajectory of causal language in early childhood remains unclear. Studies have shown that toddlers at 15 months of age can match the transitive structure of the sentence to a causative event through a mechanism known as syntactic bootstrapping (Jin & Fisher, 2014). In the current experiments, we tested whether 12-month-olds and 20-month-olds can rely on the same mechanism to acquire the association between the subject of the sentence and the causal agent of the event. Due to COVID-19 outbreak, the data collection process was interrupted. Only preliminary data was presented in this paper, with discussion focused on the implications of different potential outcomes.

Keywords: causal language, syntactic bootstrapping, word learning

Syntactic Bootstrapping as Noun Learning Mechanism in 12-Month-Olds and 20-Month-Olds Infants

Causality is central to human cognition. Without an understanding of causality, humans would find it impossible to understand the cause-and-effect nature of the world. Scientists around the world are always in search of the underlying causes: What causes the wind to blow? What causes earthquakes? What causes one plant to grow better than another plant? Non-scientists also operate both explicitly and implicitly on causality. Every time we flip a light switch, there is a causal chain: our action causes the light switch to flip, and the flipping of the light switch causes the light to turn on or off. Almost every action in our life, from typing on keyboards to baking bread, involve causality. Due to its prevalence, causality has been a topic of interest for philosophers for centuries (Hume, 1748). Over the last 40 years, developmental psychologists have also become interested in this area, and as a result a good deal is now known about the emergence of this ability in infancy. Two aspects of causality have been primarily examined: causal perception and causal reasoning. Causal perception refers to the ability automatically to perceive causal relationships between entities. This ability has been shown to exist in infants as young as 6 1⁄2 months of age. Researchers found that infants at this age can perceive causality from a direct-launching event in which one object physically causes another to move (Leslie & Keeble, 1987), and they categorize different non-causal events (i.e., one with no contact between objects before the second moves and one with a gap between two objects before one move) as equivalent and as different from a causal event (Cohen & Oakes, 1990).

The other line of research is on causal reasoning. Causal reasoning has been narrowly defined as the ability to “reason about (mostly physical) causal events” (Benton & Rakison, 2017). Compared to the early emergence of causal perception, the ability to reason with causal information emerges relatively late. Researchers have mainly focused on children beyond two years of age (Gopnik & Sobel, 2000; Gopnik, Sobel, Shulz & Glymour, 2001; Nazzi & Gopnik, 2003; Sobel, Tenenbaum & Gopnik, 2004). The focus on older children is partly due to the task demands of testing paradigms. In the classic Blicket Detector Paradigm, young children see an experimenter putting different blocks on top of a box in front of them. There are two kinds of blocks. Some blocks, when being put on the top of the box, cause the box to light up and play music. These blocks are known as “blickets”. The other kind of block would not have such an effect. In the experiment, children are asked to infer which blocks are “blickets” and which are not. Unfortunately, this paradigm is unsuitable for younger children for two reasons. First, this task requires children to have sufficient motor control to indicate which blocks are “blickets”. Second, the experimental procedure often involves verbal instruction ("See, this one set the machine off."," Now it’s your turn to make the machine go."). Younger children are likely to fail to understand the causal connotation in the sentence and would fail to comprehend the experimenter’s intention.

But how do young children begin to understand the causal meaning behind language? Most languages, like English, do not rely on explicit morphological cues to indicate causality. Although there are languages that use explicit cues for causality—for example, Kannada, a language spoken by people in southwestern India, use morphology to mark causative notion—children learning Kannada still tend to ignore the cues until age four and five (Lidz, Gleitman, &

Gleitman, 2003). Moreover, in their utterances to young children, the parents normally do not use words that explicitly refer to cause and effect. A search on the CHILDES database reveals that young children do not hear the word “cause” until 15 months of age. For children around 21 months of age, there are only thirty-six occurrences out of the 16,4518 uttered words young children heard (Bååth, 2010). In other words, causal cues are not explicit in language, and parents rarely teach children about causality in language. Therefore, for children to understand causality in language, we must assume that children learn through other aspects of linguistic input.

One promising avenue of such learning is syntax. A previous study has found a close relationship exists between verb meaning and clause structure. Fisher, Gleitman, and Gleitman (1991) found that adults systematically give verbs semantic judgment based on the syntactical context in which the verbs are embedded and vice versa. One particularly interesting observation is that when a verb appears in a transitive sentence, readers tend to interpret it as denoting causal relationships. This holds even for the verb that does not allow transitive usage. For example, “laugh” is an intransitive verb. But when being used as a transitive verb as in “Mary laughed John”, interpreters would be willing to infer this ungrammatical sentence to mean “Mary causes John to laugh”. The close connection between transitivity structure, verb, and causality is potentially crucial for causal language learning. Children pay attention to the causative notion of syntactical structure very early on. Previous studies have shown that when young children listen to a sentence with transitive structure, they tend to look longer at the scene in which two participants causally interact with each other, compared to the alternative scenario in which two participants are present but do not causally interact with each other (Jin & Fisher, 2014; Yuan, Fisher & Snedeker, 2012).

Young children's ability to use syntactical information to learn the meanings of new words is also known as "Syntactical Bootstrapping". First formalized in the 1990s, this phenomenon has been viewed as evidence for the role of syntax in verb learning (Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, 1990; Jin & Fisher, 2014; Naigles, 1990; Yuan et al., 2012; also see Fisher, Jin & Scott, 2019 for a more comprehensive review). This line of research suggests that children as young as 15 months of age, upon hearing a novel verb such as "gorping" is embedded in a transitive structure, infer that this novel verb denotes causative action. They would not make such inference if the novel verb is embedded in an intransitive structure. In the seminal study by Naigles (1990), 2-year-olds were seated on their parent's laps facing two TV-monitors. The stimuli were presented on the two monitors. First, they saw two actors engaged in a causal action while making arm gestures. They also heard "Look! The duck is gorping the bunny" accompanying the action. Then, the left and the right monitor showed different stimuli. On one monitor, the duck and the bunny performed the causal action without arm gestures. On the other monitor, the duck and the bunny did not interact, but only made arm gestures. The infants heard "Where's gorping now? Find Gorping!". They found that the toddlers looked significantly longer at the monitor showing the causal action. This was taken to be the evidence that toddlers understood the reference of "gorping" to be a causal action, and they did so by incorporating the causative notion of transitive structure.

A number of syntactical bootstrapping studies have since focused on verb learning because verbs are often considered as "hard words", that is, words in the verb class are more difficult to learn compared to words in other word classes (Gleitman et al., 2005). Indeed, converging evidence from multiple lines of research is consistent with the theory of “noun advantage” and “verb disadvantage”. Children’s early receptive and productive vocabulary is dominated by nouns, even in more “verb-friendly” language such as Korean and Chinese (Bornstein et al., 2004; Kim, McGregor & Thompson, 2000; Tardif, Gelman & Xu, 1999). Evidence also suggests that children at two years of age require more exposures for verb learning than for noun learning (Childers & Tomasello, 2006). Given the “noun advantage” and the challenges of verb learning, syntactical bootstrapping has been considered to be the core of verb learning: using the known nouns and the syntactical structure in an utterance can help to constrain and making inference about the meanings of the unknown verbs.

There is no reason to believe syntactical bootstrapping is a verb-specific learning device, however. First of all, children younger than two years of age do not yet possess a fully developed working knowledge of syntactical categories. Some have argued that very young children possess some innate knowledge of word categories. For example, Valian and colleagues argued that children’s task in learning language is mapping “the categories she possesses onto the words that she hears”. And this learning process assumes that children start learning language with abstract syntactical category already in place (Valian, Solt, & Steward, 2009). But theories like this fail to account for how children would use these “innate categories” to categorize the instances in the input they hear. (Ambridge, 2017). One study has also shown that children’s grammatical performance largely depends on the input they receive (Theakston, Lieven, Pine, & Rowland, 2001). In this study, nine children, around the age of two years, were audiotaped in the home environment for an hour every three weeks for a year. A detailed analysis of the children’s utterances and their linguistic environment showed that the children's use of syntactical frames with verbs was mainly predicted by the parents' use of syntactical frame with the specific verbs. For example, the way children use "want" in their utterances is predicted by the way their parents use "want", but it is not predicted by how the transitive verb as an abstract syntactic category is used. Therefore, an alternative account for early syntactical knowledge is needed. The usage-based theory of language learning is considered as a primary alternative for the nativist theory, and it has limited, if any, emphasis on the differences between word class (Abbot-Smith & Tomasello, 2006). Instead of learning words in terms of their word class, the usage-based approach states that language structure emerges from language use, and children acquire word meanings through a domain-general cognitive mechanism such as intention-reading and pattern findings (Brooks & Tomasello, 1999; Lieven, & Tomasello, 2008; Tomasello, 2000; Tomasello & Abbot-Smith, 2002). Therefore, even though syntactical bootstrapping contains the word "syntactical" in its name, we should be cautious when inferring the nature of the learning mechanism, not to erroneously assume that children's language learning is aided by an adult-like syntactical system that operates selectively based on word classes.

If the noun-verb discrepancy in early vocabulary does not come from the intrinsic characteristics of the word class, then it may come from the referent's different degree of abstractness and concreteness, or as Gentner and Boroditsky put it, "the degree of pre-individuation" (Gentner and Boroditsky, 2001, p.242). Noun's common referents, objects, are more "preindividuated" than verb's common referents, actions. The conceptual precursor for noun-object mapping, object individuation, is among one of the earliest mental capacities that infants possess, starting around three months of age (Aguiar & Baillargeon, 1999, 2002;

Baillargeon & DeVos, 1991; Wilcox, 1999). In contrast, the conceptual precursor for verb and motion mapping emerges relatively late. Infants' ability to detect, segment and categorize based on motion features are found to emerge during the second half of the first year and continue to develop throughout the second year. (Konishi, Pruden, Golinkoff, & Hirsh-Pasek, 2016; Pruden, Göksun, Roseberry, Hirsh−Pasek, & Golinkoff, 2012; Pulverman, Song, Hirsh−Pasek, Pruden, & Golinkoff, 2013). Furthermore, verbs' referents are often goal-directed, intentional action. The ability to perceive actions as goal-directed and intentional also emerges later than the object individuation ability (Biro & Leslie, 2007; Csibra, Gergely, Bíró, Koós & Brockbank, 1999; Myowa-Yamakoshi, Kawakita, Okanda, & Takeshita, 2011). This will also contribute to the difficulty of learning verbs. In summary, the cognitive demands in processing abstract motion event will impose a special challenge for children who are trying to learn the corresponding linguistic label and subsequently will lead to the noun-verb discrepancy.

Nevertheless, the degree of preindividuation is not a dichotomy between object and motion, noun and verb. It is a continuum. Nouns can refer to abstract concepts, which will be more difficult to acquire than verbs that refer to more concrete actions. In early studies that collected parental reports, researchers found indications that young children begin to understand verbs very early on (Benedict, 1979; Fenson et al., 1994; Goldin-Meadow, Seligman, & Gelman, 1976;). A controlled experiment using the Intermodal Preferential Looking Paradigm has shown that by 16 months of age, young infants who heard verbs would look longer at the matched scene than the unmatched scene. This was interpreted as evidence for early verb comprehension (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Together, this evidence shows that despite the “hard word” reputation, verbs are not sweepingly unlearnable in early childhood, as long as the referent of the verb is common, concrete and highly preindividuated.

The current experiments aim to address the following question: can syntactical bootstrapping be used to learn nouns? If infants already possess some knowledge of verbs, can the same principle of syntactical bootstrapping be extended to guide their noun learning? In previous syntactical bootstrapping studies, the causal scenes were composed of an agent, a patient and the causal interaction between agent and patient, and the learned component is always the mapping between the verb and causal interaction. However, both agent and patient are necessary parts of causal events, and a full understanding of causal language must comprise an understanding of the agent and the patient as well. According to the Proto-Role theory (Dowty, 1989), each verb’s argument has either Proto-Agent Entailments or Proto-Patient Entailments, or a mixture of both. Proto-Agent has properties such as causative, volitional and sentient, and Proto-Patient often has properties such as causally affected and will undergo a state of change. In English, there is a tendency to link the Proto-Agent entailments with the subject of the sentence, and Proto-Patient entailments with the object of the sentence. Such entailments warrant the semantics for the verb’s argument, and they could potentially serve as “default in the acquisition of lexical meaning” (Dowty, 1991, p. 604). Therefore, if infants already comprehend the semantics of the verb and the transitive structure that denotes the causal relationship, then they should also, based on the same principle of syntactical bootstrapping, be able to infer the mapping between the subject of the sentence to the agent of the scene.

The present experiments used a habituation paradigm to test whether 12-month-olds and 20-month-olds can learn the mapping between the subject of the sentence to the agent of the scene, as well as the mapping between the object and the patient. We selected these two younger age groups based on the following considerations. First, although infants do not start to have multi-words production until around two years of age, researchers have found that by 20 months of age infants are already sensitive to various aspects of syntactical information in the sentence (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). Second, learning the subject-agent and object-patient mappings is likely to be easier than the verb-action mapping because of the "noun advantage". More recent studies have found that infants as young as 6 months of age can already grasp the meanings of many common nouns (Bergelson & Swingley, 2012). Therefore, it is reasonable to expect that the ability to infer subject-agent and object-patient mapping predates the verb-action mapping. Third, the conceptual foundation for learning proto-agent and proto-patient entailments are already in place by the end of the first year. Infants are sensitive to the role reversal under causal context (Leslie & Keeble, 1987; Schöppner, Sodian & Pauen, 2006; Rochat, Striano & Morgan, 2004). This suggests that the semantics entailment for Proto-Agent and Proto-Patient are already available at 12 months of age, which can support the subject-agent mapping and object-patient mapping. Together, these four considerations lead us to select 12-month-olds and 20-month-olds as ages of interest. We expect this ability to continue to mature throughout the second year, and this ability will become more reliable at 20 months of age.

Because of the younger age groups, we also made several methodological changes to adapt to their cognitive capacity. First, instead of using the intermodal preferential looking paradigm (IPLP), we chose to use a habituation paradigm. At face value, these two paradigms are similar. Both use looking time as the measurement, and both include an exposure phase and a testing phase. However, the theoretical commitment between the two paradigms is different.

Habituation paradigm welcomes a process-oriented interpretation for the looking time differences (Oakes & Mandole, 2000). In the habituation paradigm, infants form and learn the mapping during the exposure phase, whereas in IPLP infants are expected to apply prior understanding to interpret the current scene. Because the goal of the current study was to investigate whether the learning is possible, rather than whether they have already learned, we consider the habituation paradigm as the more appropriate testing paradigm. Second, we reduced the complexity of visual stimuli. We used a simple direct-launching paradigm consist of two geometry shapes. Previous studies on syntactical bootstrapping have used films that rely on human actors (Naigles, 1990; Gertner, Fisher & Eisengart, 2006; Fisher, Gertner, Scott, & Yuan, 2010; Yuan et al., 2012), or shapes that have salient animacy cues (Jin & Fisher, 2014). These video clips contain a series of dynamic movements, which are more difficult to process for infants. Besides, all of them strongly invoke social interpretation. Infants may interpret the verb as denoting social interaction, rather than causality. In contrast, our design uses the classic direct-launching paradigm, which has been widely used to investigated causal perception in very young infants. (Bélanger & Desrochers, 2001; Cohen & Amsel, 1998; Leslie, 1982; Leslie & Keeble, 1987). Therefore, by using the direct-launching event as visual stimuli, we can reduce the cognitive demand for young infants, and also eliminate the potential social interpretation confounding factor.

In the current experiments, we used an infant control procedure habituation paradigm. Infants underwent two phases. In the first habituation phase, infants saw a direct launching event while listening to a spoken sentence that contained a novel label, “Neem pushes that”. Once the infants met the habituation criteria, they entered the testing phase. There were two test blocks in the testing phase. In the first block, infants saw two isolated labeling trials. A shape moved across the screen while the recorded novel label "neem" was played in the background. In one trial, the shape was the agent in the previous launching event. In the other trial, the shape was the patient. In the second block, infants were exposed to a novel sentence and a novel scene. We presented infants with two trials accompanied by the spoken sentence "That pushes neem". In one trial, infants saw the original direct-launching event. In the other trial, the causal roles in the event were reversed, with the patient shape now became the agent shape and vice versa.

If infants can learn the subject-agent mapping, then they should dishabituate more to the trial in which the patient shape is paired with label "neem", and the trial in which "That pushes neem" is paired with the original sentence. Alternatively, if they do not learn the mapping during the habituation phase, they should react equally to the two testing trials in both blocks. If infants only show the predicted novelty preference for isolated labeling block but do not show the novelty preference in the novel linguistic context block, then this may indicate that infants' early mapping ability is not yet robust enough to transfer to a novel linguistic context. However, if infants only show the predicted novelty preference for novel linguistic context block but do not show the same preference in the isolated labeling block, then this may suggest that syntactical information is necessary for infants to interpret the linguistic input.

Method

Participants

We included ten 12-month-olds (M = 12.1 months, range 11 months 19 days to 12-month 13 days; 4 females, 6 males) and seven 20-month-olds (M = 20.1 months, range 19 months 21 days to 20 months 12 days; 2 females, 5 males) participated in our study. An additional four infants were tested but not included in the final analysis due to fussiness. All were born full-term without any hearing problems or ear infections at the time of testing. Parents reported that all infants were exposed to English for the majority of the time. Parents were instructed to fill in the short form of the MacArthur CDI (Level II; Fenson et al., 2000). For 12-month-olds, the score ranged from 1 to 6 (M = 3.3; SD = 1.7). For 20-month-olds, the score ranged from 6 to 88 (M = 35.4; SD = 28.0) All infants receive either a book or a T-Shirt afterward as a gift for the participation.

Procedure

Each infant was tested individually in a dim-light room. Before the experiment, the parent was instructed to refrain from interacting with the infant but that they could stop the experiment at any time. Each infant was seated on the parent's lap in front of a TV screen. The sound was presented from two speakers in front of the television. A hidden camera located behind the TV fed the experimenter to code the looking time online. Sessions were also recorded for future offline coding to assess the reliability of the online coder.

The study was conducted with the infant control procedure on Habit 1.0. software (Cohen, Atkinson & Chaput, 2004). I used a video segment of a green ball bouncing on the screen accompanied by bell ringing as an attention-getter. During the habituation phase, each trial started playing when the infant oriented toward the screen. The trial stopped when the infant looked away for more than 1 second or looked continuously for more than 30 seconds, after which the attention-getter automatically started playing to orient the infant back to the screen. The habituation phase ended when the infant exceeded 16 trials or met the habituation criterion. We used a sliding window habituation criterion. Infants met the habituation criterion if the average looking time in three consecutive trials was lower than 50% of the average looking time in the first three consecutive trials. Once the habituation phase ended, the infant entered the test phase immediately. The test phase had two fixed-order testing blocks. Each block had two trials with order randomized. A more detailed description of the test material was described in the section below.

Before the habituation phase and after the test phase, the infant received two attention-checker trials. The infant saw the same silent bright-colored cartoon video clip for the pre-experiment trial and the post-experiment trial. The purpose of these two attention-checker trials was to measure the infant's baseline attention. An attentive infant should look equally long during the trial before the experiment and the trial after the experiment. In contrast, if the infant looked significantly shorter in the post-experiment trial than the pre-experiment trial, then it was possible that the drop in looking time during the habituation phase was due to fatigue rather than the effect of learning.

Stimuli

Habituation stimuli. During the habituation phase, the infant was habituated to a simple Michottian launching event accompanied by an English sentence ("Neem pushes that") spoken by a female native speaker (See figure 1 for a schematic illustration of the stimuli). The Michottian launching event was created with Macromedia Director 8.0 for PC.

In the Michottian launching event, the infant saw a blue circle with a red heart (patient) staying still at the center of the stage. A black square with a yellow star (agent) moved from outside of the screen into the stage until making contact with the blue circle. Upon the contact, the black square stopped moving and the blue circle started moving horizontally along the same direction the black circle was moving. The blue continued moving until it was outside of the screen and no longer visible to the viewer. A screen dropped and lifted. The same sequence repeated until the trial ended. Each infant saw the shapes moving from left to right as well as the right to left. The moving direction was counterbalanced across participants.

The linguistic stimuli were adapted to the infants' linguistic capacity. Since I was interested in whether the infant could incorporate the syntactic and semantic information in the sentence to infer the mapping between the subject of the sentence and the agent of the scene, I selected "pushes" and "that" as two words the infant already knew. "Push" was selected as the known verb because it is among the early verbs vocabulary that infants acquire (Benedict, 1979; Fenson et al., 1994; Goldin-Meadow et al. ,1976;). "That" was selected as the known word because this pronoun has the highest frequency in young children's early linguistic input (Bååth, 2010). "Neem" was the novel label given to the infants.

The video and the audio were synchronized. The infant heard "Neem" upon the black square entering into the view. Around the moment when the two shapes making contact, the infant heard "pushes". Infant heard "that" as the blue circle departed.

Test stimuli. There were two testing blocks: the label-only block and the novel-sentence block. Each testing block contained two trials. Each trial began as the infant oriented toward the screen and ended as the infant turned away for more than one second.

During the label-only block, in one trial, the infant saw the black square (the agent shape) moved similarly across the screen. In the other trial, the blue circle (the patient shape) moved across the screen. As the shape passing through the screen, the label "Neem" spoken by the same female native English speaker was repeated three times throughout the audio. The order in which the infant saw agent shape or patient shape passing through was counterbalanced across participants.

During the novel-sentence block, in one trial, the infant saw the same direct launching event from the habituation phase. In the other trial, the infant saw the direct launching event in the same direction, but with the role of agent and patient reversed. Now, the agent shape became the patient shape, and the patient shape became the agent shape. A sentence spoken by the same female native English speaker was synchronized to the video in the same way during the habituation phase. This time, the sentence provided was “That pushes neem”. The order of the two trials was also counterbalanced.

Prediction and Preliminary Results

If the infant can obtain a mapping between the subject of the sentence and the agent of the event, one should look longer at the trials with incongruent mapping relationship. If the mapping is strong enough, then the mapping relationship should also hold in the novel linguistic context (i.e., when the sentence is new). Therefore, In the label-only block, the infant should look longer at the trial in which the patient block was paired with the label “Neem”. In the novel-sentence block, the infant should look longer at the trial in which the direct launching event is identical, but the sentence had the subject and the object reversed. However, it is also likely to be the case that the association between the label and the shape is relatively weak, such that the infant could only detect the incongruence in the label-only block. In this case, we predict that the infant would only look longer at the patient-label pairing trial during the label-only block. In the novel-sentence block, they either look longer at the trial with reversed launching event but congruent pairing or look equally long across the two trials.

Unfortunately, the data collection process was interrupted due to the outbreak of COVID-19. The current sample size is insufficient to perform meaningful statistics analysis. I will only present descriptive statistics for qualitative comparison. On average, 20-month-old infants looked longer (M = 22.09s, SD = 11.01) at the test stimuli than 12-month-old infants (M = 14.95s, SD = 9.05). Twelve-month-old infants (N= 10) looked longer in the trials with inconsistent pairing than in the consistent pairing during both the label-only block (Inconsistent: M = 15.93s, SD = 8.80; Consistent: M = 13.47s, SD = 9.27) and the novel-sentence block (Inconsistent: M = 16.71s, SD = 8.69; Consistent: M = 13.70s, SD = 10.38). 20-month-old infants (N = 7) showed a reversed trend, looking longer in the trials with consistent pairing than in the inconsistent pairing during both the label-only block (Inconsistent: M = 18.69s, SD = 12.07; Consistent: 25.80s, SD = 10.85) and the novel-sentence block (Inconsistent: M = 20.41s, SD = 12.45; Consistent: 23.47s, SD = 9.60). I plan to fit a mixed-effect linear regression model with age, sex, block, pairing-type and the interaction between block and pairing-type as fixed-effect variables and individual participants as random effects. I will also evaluate the model-fit with permutation test and Bayesian Factor analysis.

Discussion

At the current stage, our limited data prohibits any conclusion to be made. Therefore, the discussion section of this paper will be devoted to hypothetical scenarios. I will first focus on the major predictors included in the mixed-effect model planned to be used. The response is looking time duration as measured by the online coder using Habit software. The predictors planned to be included are Pairing-Type (categorical variable, consistent pairing or inconsistent pairing), Block (categorical variable, label-only block or new-sentence block) and Age (categorical variable, 12-

month-olds or 20-month-olds.). I will first discuss the potential outcomes of these predictors with possible interactions. Following the discussion on these variables, I will elaborate on their implications if those effects were to be found. I will also make suggestions for future research directions.

My hypothesis predicts that Pairing-Type should be a major predictor for infants' looking time duration in the test phase. If the infants have acquired the mapping between the subject of the sentence and the agent of the scenes, they should look longer at the trials in which the mapping is inconsistent with what they have been habituated too. Therefore, inconsistent trials, in which the linguistic label was paired with a different object in the event, should lead to a longer duration than consistent trials, in which the pairing was retained.

If we do observe this main effect of Pairing-Type, then this would add to our understanding of early language acquisition critically. My current findings constitute new evidence for the early onset of using sentence structure to guide noun learning. In previous word learning studies with young infants, nouns are often presented in isolation, or in a way not necessarily informed by the syntactical structure of the sentence ("Do you see the X?") (Bergelson & Swingley, 2012, 2013). Similarly, studies on infants' ability to associate nouns with categories also presented nouns without meaningful syntactical constraints (Waxman & Markow, 1995; Waxman & Booth, 2003; Waxman & Braun, 2005). This words-in-isolation approach is not widely used with older children, however. Word learning studies that presented participants with grammatical sentences often recruited participants who are around and beyond two years of age (Brown, 1957; Naigles, 1990; Naigles & Kako, 1993; Soja, Carey & Spelke, 1991; Soja, 1992). One rare exception is Jin and Fisher's work in 2014, showing that the usage of syntactical information can be detected as early as 15-month-olds (Jin & Fisher, 2014). But still, there is a gap in the literature between early competency for noun-in-isolation and late competency for noun-in-construction that calls for a more careful exploration of the possible mechanisms in which young infants can acquire the semantics of the noun. It is reasonable to hypothesize that syntax helps early on. At 12 months of age, infants can represent multilevel statistical regularities, establish form-based category categorization from and use word orders as cues to distinguish between sentence types (Geffen & Mintz, 2015; Gómez, & Lakusta, 2004; Saffran & Wilson, 2003). These abilities are considered as important for syntax learning. With their presence at an early age, combined with infants' early competency of word meanings, young infants should be equipped with the necessary conditions to incorporate their syntactical knowledge into determining word meanings. Yet, there is still no empirical evidence supporting this prediction. Therefore, if we observe the desired effect, the evidence should address this gap by showing that around one year, infants can utilize the transitive structure to attribute agency to the novel word appearing in the subject argument of the sentence.

Alternatively, Pairing-Type may end up being an insignificant predictor in the model. Here, I will consider the three most likely alternative outcomes: a) No main effect of Pairing-Type, but an interaction effect between Pairing-Type and the Block. b) No main effect of Pairing-Type, but an interaction effect between Pairing-Type and the Age. c) No main effect of Pairing-Type, and no interaction effect between Pairing-Type or any other variable. This is certainly not an exhaustive list of all possible interactions. They are selected because there are the most plausible, most relevant and most interesting potential outcomes based on the current theorizing.

No Main effect of Pairing-Type, But an Interaction Effect Between Pairing-Type and the Block

There are two directions for interactions: the Label-Block-Driven scenario and the Sentence-Block-Driven scenario. In the Label-Block-Driven Scenario, the infants looked longer at the inconsistent trial than the consistent trial in the Label-Only Block but failed to do so in the Sentence-Only-Block. Vice versa for the Sentence-Block-Driven scenario.

Label-Block-Driven scenario. The infants looked longer at the inconsistent trial than the consistent trial in the Label-Only Block, but they failed to do so in the Sentence-Only-Block. This finding suggests that infants successfully associated the novel word “neem” and the agent of the event when presented in isolation. But, infants failed to generalize their association in a novel linguistic context. This outcome can be interpreted as word order playing a crucial role in encoding the association between the subject and the agent. Word order is an early and salient cue for syntactical information. The ability to encode word order has been found in infants as young as two-month-olds (Mandel, Nelson & Jusczyk, 1996). And later on, around two years of age English-Speaking children can generalize their knowledge of word-order to novel verbs (Akhtar, 1999; Gertner, Fisher & Eisengart, 2006). When the stimuli presented in the Sentence block has conflicting word order with the habituation trials, the infants, prioritizing the word-order information, would fail to retrieve the association between the original-subject and the original-agent. In contrast, the label-only block does not contain such conflicting information. Infants can still retrieve the previously acquired associations.

Another interpretation of the differences in the agency bias in the event and subject’s privileged status in the linguistic input. Agency bias refers to the finding that in a causal event, infants tend to prioritize encoding the agent over the recipient. Cohen and Oakes (1993) demonstrated that when habituated to a causal-event, 10-month-olds infants notice a change in the agent but failed to notice a change in recipient. This agency bias has also been replicated in causality involving social agents. Twelve-month-old infants were shown selectively to pay more attention to the social agent and their actions but less so to the objects being acted on (Moore, 1999). A parallel bias is also present in linguistic input. The subject enjoys a privileged status in sentences across languages. Among all the studied languages with dominant word order, 96.3% of the languages require subject preceding objects (Dryer, 2005). When asked to represent an event through gestures, native speakers of languages with different word orders uniformly choose to represent the subject before the object, suggesting that subject first and object latter constitutes a natural sequencing of the event (Goldin-Meadow, So, Özyürek, & Mylander, 2008). One potential explanation of this privileged status is the cognitive prominence of the agent (Tabullo et al., 2012). For a language-learning infant, the agency bias and the privileged status of the subject may jointly contribute to selectively encoding subject-agent pair. Switching the order of the agent and the patient and the order of the subject and the object could significantly interfere with the infants' representations of the association, causing them to show different looking patterns in the two blocks.

Finally, the pattern can also be explained from a methodological perspective. The performance in the sentence-block was primarily driven by Thompson-Spencer dishabituation, rather than a failure to learn the content of our stimulus. Thompson-Spencer dishabituation is defined as a dishabituated response to an already habituated stimulus, after the presence of the novel stimulus (Thompson & Spencer, 1966). Kaplan and Werner conducted a series of experiments showing that infants would dishabituate to the old stimulus after showing the Thompson-Spencer effect (Kaplan & Werner, 1986, 1987). After the infants have habituated to stimulus A, they were presented with stimulus B, then A again. Infants not only showed a novelty response to B but also to A upon seeing them again. Kaplan and Werner argued that Thompson-Spencer dishabituation cannot be simply attributed to a decaying representation of A. Rather, this shows that novel stimulus can influence the infants' arousal state. Seeing novel stimulus B "excited" the infants, leading them to look longer at the already-habituated stimulus A as well. In the current experiment, the Label-Only block preceded the Sentence-Only block. Compared to the Sentence-Only block, Label-Only block was also more perceptually different from the stimulus presented in the habituation phase. Therefore, the Label-Only block can be construed as a novel stimulus, and Sentence-Only block as an approximate of the habituated stimulus. Thus, in Sentence-Only block the infants were experiencing Thompson-Spencer dishabituation. Note that this interpretation does not imply that infants have failed to retain or retrieve the associations formed in the habituation phase. Instead, Thompson-Spencer dishabituation interfered with the looking time behaviors. It made the learned effect, if there is any, difficult to be detected based on differential looking time.

Sentence-Block-Driven scenario. The other possible outcome with interaction effect between Block and Pairing-Type is that infants would look longer at the inconsistent trials in the Sentence-Only block, but it would show an ambiguous looking pattern in the Block-only trial. This interaction would suggest that infants can generalize their associations between the novel noun and the agent of the scene to a novel linguistic context. However, when the novel noun is presented without any linguistic context, they fail to retrieve the representation.

One possible interpretation is the task demand. Our stimulus may be too complex for 12-month-olds while being too simple for 20-month-olds. According to the multi-factor model proposed by Hunter and Ames (1988), infants' visual preferences for novelty or familiarity are moderated by three factors, age, task complexity and familiarization time. In our case, some 12-month-olds may find the stimulus during the habituation phase overwhelmingly complex. Their longer looking time indicates familiarity, rather than novelty. In contrast, 20-month-olds may find the stimulus simple and have successfully acquired the correspondences between the agent of the scene and the subject of the sentence. Yet, they may choose to look longer at the scenes that match their understanding. This interpretation of longer looking time is adopted by Intermodal Preferential Looking Paradigm (IPLP) devised by Golinkoff and colleagues and widely accepted in language acquisition literature (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). Although both IPLP and habituation relies on looking time differences as a measurement for learning outcome, they lead to opposite interpretation. At the current stage, it remains unclear how these looking time paradigms connect. But it is at least in principle possible that 20-month-olds chose to look longer at the consistent trials because it matched what they have learned from previous habituation trials.

A perhaps more intriguing possibility is that the performance in the Label-Only trial was caused by not able to retrieve the association they have acquired during the habituation phase. This suggests that syntactical support is also crucial for acquiring noun meaning. In an early study using the habituation paradigm, Casasola and Cohen (2000) found that 14-month-olds failed to form associations between words and actions. They attributed such failure to young infants' underdeveloped information processing capacities unable to process linguistic input and the causal event simultaneously. In contrast, the 15-month-olds infants in Jin and Fisher's (2015) study were competent to process inputs from two domains, and they were capable to parse the sentence structure for constraining verb meanings. While the infants may have rapidly matured over one month, a more likely explanation was the difference in the linguistic input. In Casasola and Cohen (2000), the linguistic input was an isolated novel word label, including "neem" and "lif". In Jin and Fisher (2015), however, the novel words were all embedded in the transitive structure during the learning phase. The additional help from the syntactical structure may have contributed to the young infants' fast learning. Similarly, in the current study's Label-Only block, when the target word was deprived of the syntactical structure, infants may have a more difficult time retaining and retrieve the association, whereas, in the Novel-Sentence block, the presence of a syntactical structure further scaffolded the association.

No main effect of Pairing-Type, but an Interaction Effect Between Pairing-Type and the Age

20-month-olds may show a greater discrepancy between the inconsistent and the consistent trials than 12-month-olds. This is a highly possible scenario. English speaking children normally do not start to produce multiple-word utterances toward the latter half of the second years (Valian, 1986). And when they do so, it is questionable whether the usage of multiple-word utterance suggests the underlying grammatical competence (Tomasello, 2000). Therefore, 20-month-old's superior performance in the current task might be a natural consequence of their more advanced language skills. Alternatively, it is also possible that the results favor 12-month-olds over 20-month-olds. This reversed trend is likely due to the method used. Our current preliminary data suggests that the older age group has a higher dropout rate compared to 12-month-olds. 20-month-olds toddlers were more likely to be bored with the stimuli, and their advanced motor skills enabled them to depart from the study on their own. In future studies, a more age-appropriate design should be used to test the linguistic competency of 20-month-olds.

No main effect of Pairing-Type, and No Interaction Effect Between Pairing-Type or Any Other Variable

While this is certainly a disappointing outcome, it is also an important opportunity to once again scrutinize the differences between the current study and previous work. In our current experiment, two major factors might attribute to the failure of the null hypothesis. To start with, the nature of the visual stimuli. In our current study, we used a simple direct-launching event. The agent and the patient involved in the causal action were geometrical shapes marked in different colors and shapes. In contrast, earlier syntactical bootstrapping studies used video clips of real human actors (Naigles, 1990; Naigles & Kako, 1993). When animations of geometrical shapes were being used, such as the case in Jin and Fisher's (2014) experiments, both the agents and patients were depicted as having schematic eyes. Eyes are strong cues for animacy (Rakison & Poulin-Dubois, 2001). Young children in previous studies may have been aided by the strong cues for animacy in identifying agents and patients. On the language side, it has been shown that infants at 12-month-olds recognize speech’s social-communicative role (Martin, Onishi, & Voulumanous, 2012). This is consistent with the recent theories in language development, which regard social reasoning as playing a pervasive role in language development (Bohn & Frank, 2019). Consequently, young infants may be more motivated to connect linguistic input with a social scene than the non-social abstract representation of causality in our experiment. Future research can systematically manipulate and compare the social factors and see to what degree social factors can help infants parse the linguistic structure, the event, and represent the associations between the two.

In addition to the visual stimulus, our linguistic stimulus also differs. The goal of the current experiment was to see if syntactical bootstrapping can be generalized to learning nouns. While failure to reject the null hypothesis does not imply the null hypothesis is correct, it is still possible that the answer to that question is “no”. According to Gentner and Boroditsky’s natural partition hypotheses, nouns are learned before verbs because the references of the nouns are highly preindividuated and readily available in the perceptual input (Gentner & Borodistky, 2001; Gentner, 2006). As a result, noun learning does not rely heavily on linguistic information as verbs do. Nevertheless, this is not to suggest that nouns learning do not rely on linguistic information. As young as 19 months of age, infants can infer the meanings of the novel nouns based on some familiar verbs (Ferguson, Graf, & Waxman, 2014). And by 24 months of age, they can extract the meanings of the novel nouns from listening to linguist input alone (Ferguson et al., 2017). However, in the current case, the highly individuated nature of the agent and the patient in our visual stimulus may have undermined parsing and utilizing the linguistic structure. Therefore, future research can potentially test on whether using a less individuated event, such as a change in state or spatial position, may lead infants to rely more heavily on the linguistic structure.

It is impossible to exhaust all the possible outcomes in the current study. Nevertheless, the ones enumerated above lead to some interesting questions that can be potentially pursued in the future. First of all, the methodological issues involved in testing infants. While there have been multiple reviews on infant paradigms that rely on looking time as measurement, which including the habituation paradigm and the Intermodal Preferential Looking Paradigm (for habituation, see Bornstein, 1985; Colombo & Mitchell, 2009; Turk-Browne, Scholl & Chun, 2008; Kavšek, 2013; For IPLP, see Golinkoff et al., 2013; Hirsh-Pasek & Golinkoff, 1996). Unfortunately, one thing that still needs more clarity is the connection and disconnection between these two paradigms. This is particularly problematic because the same observation ("longer looking time to X") in these two paradigms can lead to opposite interpretations. In our experiment, the longer looking time signals that infants detect a mismatch between what they have learned in the habituation phase. In a typical IPLP, it signals that infants "understand" the utterances, and are looking for the matching scenarios. Given how prevalence these two methods are in developmental research, a pressing research question to be addressed is the scope and limitation of each of the paradigm. A rigorous method is a prerequisite for high-quality data. And it is the necessary foundation for knowing the underlying mechanisms, and for robust theory building.

In summary, if the results in the future are consistent with my hypothesis, then they would constitute the evidence showing that the sensitivity to syntactic structure plays a critical role even before infants start to use multiword utterance. Furthermore, this also suggests that syntactic bootstrapping is not necessarily a verb-specific learning mechanism. Infants can also use their syntactic knowledge to constrain the meanings of nouns in the sentence. I have also discussed some interesting implications if the results were to come out inconsistent with my hypothesis. First, there might be an interaction effect between the Pairing-Type and the Block-Type. Depending on the different potential scenarios, the results are going to provide important insights into language learning from three different perspectives: whether word order is crucial for early noun learning, whether infants incorporate agency-bias in forming the association, and whether syntactic cue is necessary for early noun learning. Furthermore, the results can also highlight some methodological issues involved in the usage of looking time paradigm in language-learning studies: whether Thompson-Spencer dishabituation and the demanding task complexity are responsible for infants' looking behavior, and therefore undermine the measurement's validity.

References

Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. Cognitive psychology, 39(2), 116-157.

Aguiar, A., & Baillargeon, R. (2002). Developments in young infants' reasoning about occluded objects. Cognitive psychology, 45(2), 267-336.

Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. Journal of child language, 26(2), 339-356.

Ambridge, B. (2017). Syntactic Categories in Child Language Acquisition: Innate, Induced, or Illusory?. In Handbook of categorization in cognitive science (pp. 567-580). Elsevier.

Bååth, R. (2010). ChildFreq: An online tool to explore word frequencies in child language. Lucs Minor, 16, 1-6.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. Child development, 62(6), 1227-1246.

Bélanger, N., & Desrochers, S. (2001). Can 6-month-old infants process causality in different types of causal events?. British Journal Of Developmental Psychology, 19(1), 11-21. doi: 10.1348/026151001165930

Benedict, H. (1979). Early lexical development: Comprehension and production. Journal of Child Language, 6(2), 183-200. doi:10.1017/S0305000900002245

Benton, D.T., & Rakison, D.H (2017). Piecing the World Together One Cause at a Time: Causal discovery in infants and young children. Unpublished Manuscript.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. Proceedings Of The National Academy Of Sciences, 109(9), 3253-3258.

doi:10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. Cognition, 127(3), 391-397.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal−directed actions: development through cue−based bootstrapping. Developmental science, 10(3), 379-398.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. Cognition, 56(1), 1-30.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. Annual Review of Developmental Psychology, 1.

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., ... & Vyt, A. (2004). Cross−linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. Child development, 75(4), 1115-1139.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. Language, 720-738.

Brown, R. W. (1957). Linguistic determinism and the part of speech. The Journal of Abnormal and Social Psychology, 55(1), 1.

Casasola, M., & Cohen, L. B. (2000). Infants' association of linguistic labels with causal actions. Developmental Psychology, 36(2), 155.

Childers, J. B., & Tomasello, M. (2006). 12 Are Nouns Easier to Learn Than Verbs? Three Experimental Studies. Action meets word: How children learn verbs, 311.

Cohen, L., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. Infant Behavior And Development, 21(4), 713-731. doi:10.1016/s0163-6383(98)90040-6

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. Developmental Psychology, 29(3), 421.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. Neurobiology of learning and memory, 92(2), 225-234.

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason'in infancy. Cognition, 72(3), 237-267.

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. Language, 67(3), 547-619. doi:10.2307/415037

Dryer, M. (2005). Order of subject, object and verb. The world atlas of language structures, 330-333.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., Pethick, S., . . . Stiles, J. (1994). Variability in Early Communicative Development. Monographs of the Society for Research in Child Development, 59(5), I-185. doi:10.2307/1166093

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15-and 19-month-olds. Cognition, 131(1), 139-146.

Ferguson, B., Graf, E., & Waxman, S. R. (2018). When veps cry: two-year-olds efficiently learn novel words from linguistic contexts alone. Language Learning and Development, 14(1), 1-12.

Fisher, C. , Gertner, Y. , Scott, R. M. and Yuan, S. (2010), Syntactic bootstrapping. WIREs Cogni Sci, 1: 143-149. doi:10.1002/wcs.17

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. Cognitive psychology, 23(3), 331-392.

Fisher, C., Jin, K. S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. Topics in cognitive science.

Geffen, S., & Mintz, T. H. (2015). Can you believe it? 12-Month-olds use word order to distinguish between declaratives and polar interrogatives. Language Learning and Development, 11(3), 270-284.

Gentner, D. (2006). Why verbs are hard to learn. Action meets word: How children learn verbs, 544-564.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. Language acquisition and conceptual development, 3, 215-256.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. Psychological science, 17(8), 684-691.

Gleitman, L. (1990). The structural sources of verb meanings. Language acquisition, 1(1), 3-55.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. Language learning and development, 1(1), 23-64.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. Cognition, 4(2), 189-202.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. Proceedings of the National Academy of Sciences, 105(27), 9163-9168.

Golinkoff, R., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. Journal of Child Language, 14(1), 23-45. doi:

10.1017/S030500090001271X

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned?. Perspectives on Psychological Science, 8(3), 316-339.

Gómez, R. L., & Lakusta, L. (2004). A first step in form−based category abstraction by 12−month−old infants. Developmental science, 7(5), 567-580.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. Child development, 71(5), 1205-1222.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. Developmental psychology, 37(5), 620.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension.

Hume, D. (1748). An enquiry concerning human understanding.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. Advances in infancy research.

Jin, K. S., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. In Boston University conference on language development. Boston, MA: Cascadilla Press. Retrieved from http://www. bu. edu/bucld/files/2014/04/jin. pdf.

Kaplan, P. S., & Werner, J. S. (1986). Habituation, response to novelty, and dishabituation in human infants: Tests of a dual-process theory of visual attention. Journal of experimental child psychology, 42(2), 199-217.

Kaplan, P. S., & Werner, J. S. (1987). Sensitization and dishabituation of infant visual fixation. Infant Behavior and Development, 10(2), 183-197.

Kavšek, M. (2013). The comparator model of infant visual habituation and dishabituation: Recent insights. Developmental Psychobiology, 55(8), 793-808.

Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. Journal of Child Language, 27(2), 225-254.

Konishi, H., Pruden, S. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Categorization of dynamic realistic motion events: Infants form categories of path before manner. Journal of experimental child psychology, 152, 54-70.

Leslie, A. M. (1982). The Perception of Causality in Infants. Perception, 11(2), 173–186. https://doi.org/10.1068/p110173

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. Cognition, 25(3), 265-288.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. Cognition, 87(3), 151-178.

Lieven, E., & Tomasello, M. (2008). Children's first language acquistion from a usage-based perspective. In Handbook of cognitive linguistics and second language acquisition (pp. 178-206). Routledge.

Mandel, D. R., Nelson, D. G. K., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. Cognitive Development, 11(2), 181-196.

Martin, A., Onishi, K. H., & Vouloumanos, A. (2012). Understanding the abstract role of speech in communication at 12 months. Cognition, 123(1), 50-60.

Moore, C. (1999). Intentional relations and triadic interactions. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), Developing theories of intention: Social understanding and self-control (p. 43–61). Lawrence Erlbaum Associates Publishers.

Myowa-Yamakoshi, M., Kawakita, Y., Okanda, M., & Takeshita, H. (2011). Visual experience influences 12-month-old infants' perception of goal-directed actions of others. Developmental psychology, 47(4), 1042.

Naigles, L. (1990). Children use syntax to learn verb meanings. Journal of child language, 17(2), 357-374.

Naigles, L. G., & Kako, E. T. (1993). First contact in verb acquisition: Defining a role for syntax. Child development, 64(6), 1665-1687.

Nazzi, T., & Gopnik, A. (2003). Sorting and acting with objects in early childhood: An exploration of the use of causal cues. Cognitive Development, 18(3), 299-317.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. Cognitive Development, 5(2), 193-207.

Oakes, L., & Madole, K. (2000). The Future of Infant Categorization Research: A Process-Oriented Approach. Child Development, 71(1), 119-126. doi:10.1111/1467-8624.00125

Pruden, S. M., Göksun, T., Roseberry, S., Hirsh−Pasek, K., & Golinkoff, R. M. (2012). Find your manners: How do infants detect the invariant manner of motion in dynamic events?. Child development, 83(3), 977-991.

Pulverman, R., Song, L., Hirsh−Pasek, K., Pruden, S. M., & Golinkoff, R. M. (2013). Preverbal infants' attention to manner and path: Foundations for learning relational terms. Child Development, 84(1), 241-252.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. Psychological bulletin, 127(2), 209.

Rochat, P., Striano, T., & Morgan, R. (2004). Who is doing what to whom? Young infants' developing sense of social causality in animated displays. Perception, 33(3), 355-369.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12−month−old infants. Infancy, 4(2), 273-284.

Schöppner, B., Sodian, B., & Pauen, S. (2006). Encoding action roles in meaningful social interaction in the first year of life. Infancy, 9(3), 289-311.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's causal inferences from indirect evidence: Backwards blocking and Bayesian reasoning in preschoolers. Cognitive science, 28(3), 303-333.

Soja, N. N. (1992). Inferences about the meanings of nouns: The relationship between perception and syntax. Cognitive development, 7(1), 29-45.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. Cognition, 38(2), 179-211.

Tabullo, Á., Arismendi, M., Wainselboim, A., Primero, G., Vernis, S., Segura, E., ... & Yorio, A. (2012). On the learnability of frequent and infrequent word orders: An artificial language learning study. The Quarterly Journal of Experimental Psychology, 65(9), 1848-1863.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of English and Mandarin. Child Development, 70(3), 620-635.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. Journal of child language, 28(1), 127-152.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. Psychological review, 73(1), 16.

Tomasello, M. (2000). Do young children have adult syntactic competence?. Cognition, 74(3), 209-253.

Tomasello, M., & Abbot-Smith, K. (2002). A tale of two theories: Response to Fisher. Cognition, 83(2), 207-214.

Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. Frontiers in human neuroscience, 2, 16.

Valian, V. (1986). Syntactic categories in the speech of young children. Developmental psychology, 22(4), 562.

Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited-scope formulae? The case of children's determiners. Journal of child language, 36(4), 743-778.

Waxman, S., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11−month−olds. Developmental Science, 6(2), 128-135.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. Cognition, 95(3), B59-B68.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-to 13-month-old infants. Cognitive psychology, 29, 257-302.

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. Cognition, 72(2), 125-166.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. Child development, 83(4), 1382-1399.

Appendix

Figure 1

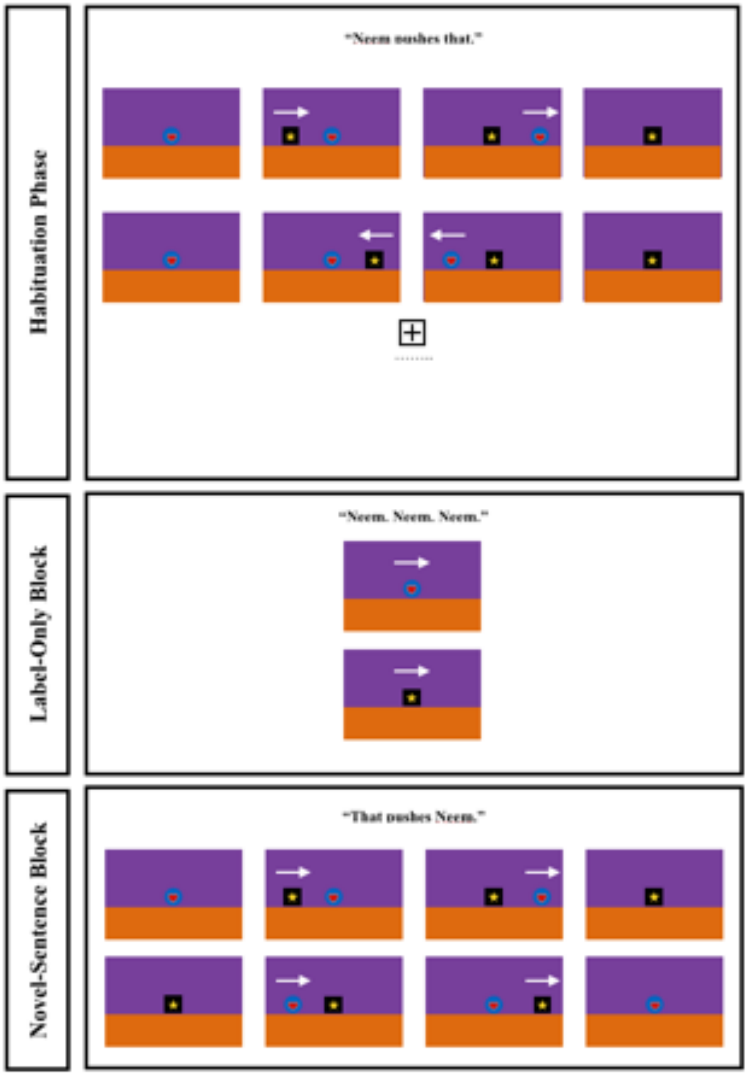


Figure 1A Schematic illustration of the testing procedure and the stimuli .

Part II. Theoretical Review

Theoretical Foundations of the Infant Looking Time Paradigms: Familiarization, Habituation and Violation of Expectation

Abstract

In the past decades, the looking time measurement has been the backbone of developmental psychology. Most of our understanding of infant perceptual and cognitive development came from research using at least one of the three looking time paradigms: habituation, familiarization, and Violation of Expectation. However, the myriad claims supported by infant looking time paradigms suffer from what Richard Aslin called a “many-to-one” problem: there are many different postulated hypotheses, but only one measurement available (Aslin, 2007). The vast underdetermination between the evidence and the interpretation originates from the lack of attention devoted to the theoretical foundations of the looking time paradigms. In this paper, I surveyed and compared the four most prevalent theories in the field: The Comparator Theory, the Multifactor Model, the Object File Theory, and the Dual Process Theory. I analyzed each theory’s strengths and weaknesses in the context of experimental design and data interpretation. I also compared the four theories against each other to assess their explanatory scope. I arrived at the conclusion that none of the theories currently available is sufficient to justify the connection between the evidence and the interpretation. In the future, more systematic investigations are needed to construct a more precise, quantitative interpretation framework to guide empirical research.

Keyword: infant looking time; method; theory

Theoretical Foundations of the Infant Looking Time Paradigms: Familiarization, Habituation and Violation of Expectation

In the realm of psychology, few terms are as loaded as "habituation". This term refers to a learning mechanism that loosely covers the behavior characterized as reducing a response toward a repeated stimulus over a short period. This phenomenon is so prevalent among many different organisms that observations were made about it centuries before the formation of modern science. In a short historical survey, neuroscientist Richard Thompson pointed out that there were even observations about "habituation" in Aesop's Fable: a fox was frightened when it first saw a lion, not so frightened for the second time, and at the third time the fox was even bold enough to have a conversation with the lion (Thompson, 2009). This interest in habituation was transformed into experimental investigations in the early 20th century. Since then, a burst of experiments approached this issue in different organisms, ranging from single-cell organisms to human infants. More and more, scientists have started to recognize that habituation is probably one of the most fundamental learning mechanisms among living things.

Habituation as a learning mechanism especially attracts the attention of developmental psychologists. "How do infants learn" and "What do infants learn" remain two core questions in the field. But these questions are difficult to answer due to infants' limited behavioral repertoire: young infants do not follow instructions nor give verbal feedback. Since Fantz (1961) first discovered that young infants would have different fixation times toward different visual patterns, looking time has become an important behavioral measurement to reveal the mental capacity of preverbal infants. Researchers saw great opportunities in this measurement and began to combine the fixation duration with habituation. They recognized that infants' looking time patterns might be good indications of the underlying perceptual and cognitive processes. Later, three related looking time paradigms, *familiarization*, *habituation*, and *violation of expectations,* were established and are still widely being used by researchers today.

On the one hand, these paradigms look similar at face value: all three are looking time paradigms that use infants' eye gaze duration as measurement, and all three contain at least two distinctive phases: a pretest phase and a test phase. On the other hand, however, researchers using each paradigm tend to bring in different assumptions about infants' perception and cognition. Even if all three paradigms measure infants' looking time, the interpretations of the data diverge significantly. As a result, the true limits of infants' perceptual and conceptual capacities remain an ongoing debate. In the following paragraphs, I will first briefly describe a specific paradigmatic experiment that uses each paradigm, and then summarize the systematic differences among the three paradigms.

Example: Familiarization

In the pretest phase, each infant was presented with six trials. In each trial, the infant saw two randomly selected pictures of cats or dogs presented side by side for 15s. Once the six 15s trials were over, each infant was presented with two test trials. Each test trial lasts 10s, and infants were presented with two pictures side by side. In the test trials, one picture was a novel exemplar from the familiar category (a picture of a new cat or new dog), and another picture was a novel exemplar from a novel category (a picture of the bird). Infants' looking time during the pretest phase and the testing phase was measured and compare

The procedure used in Quinn, Eismas & Rosenkrantz (1993), Experiment 1

Example: Habituation

Infants were assigned to two groups, the direct launching group or the delayed reaction group. In the pretest phase, depending on the group assignment, each infant was presented with a film segment depicting either a direct launching causal event or a delayed reaction event. Each trial began when the infant directed his or her eye gaze toward the screen, prompted by flashing light. Each trial ended when the infant looked away for more than 1 second. The total length of the pretest phase depended on the infant' s looking time. When the infant's mean-looking time in three consecutive trials was 0.5 seconds less than the mean-looking time in the first three consecutive trials, the pretest phase ended. The minimum number of pretest trials was 6, and the maximum number of pretest trials was 18. After the pretest phase and a 40s break followed immediately after, each infant was shown the animated segment reversed. The looking time during the pretest and the testing phase was measured and compared.

The procedure used in Leslie & Keeble (1987), Experiment 1

Example: Violation of Expectation

Infants were seated in front of the experimental apparatus, which was capable of showing two types of physical events: the impossible event and the possible event. In the impossible event, the flat screen would fall flat, as if it fell through the block behind it. In the possible event, the flat screen would be stopped by the block in the middle. The event used during the pretest trial was identical to the impossible event, except that there were no blocks behind the screen. During the pretest phase, each trial ended when either the infant looked away for 2 consecutive seconds or the infant had been looking at the trial for more than 120 seconds. After each trial, there was a 3s inter-trial interval, and the event repeated itself. The pretest phase ended when the infants' mean looking time in three consecutive trials had a 50% or more decrease compared to the mean-looking time in the first three trials. The maximum number of pretest trials was 14. After the pretest trial, infants were shown the woodblock standing clearly behind the screen twice, each time for 3s. Then, the test phase began. Infants were shown 3 pairs of test events, with the possible event and impossible event presented alternatively. Another group of participants was assigned to "the control experiment", in which the testing procedures remained identical, except for the inter-phase events. In the control experiments, the infants were shown the woodblock standing next to the screen, rather than behind the screen. In both experiment Infants' looking time during the pretest trials and testing trials was measured and compared.

The procedure used in Baillargeon, Spelke & Wasserman (1985)

In summary, in all three paradigms the pretest phase consists of multiple trials. Depending on the research topic and the paradigm used, the number of trials can range from as few as four to as many as twenty. The content and display of the trial also vary depending on the research topic. Some researchers use static images, while others use animated event segments on the screen. Some researchers use a hand-crafted apparatus to present a scene in a real-world display to test physical reasoning while others use computer animations on the screen to test moral understanding. The diversity of the testing procedures is shown in the preliminary summary table (Table 1, Figure 1).

In addition to the implementational differences, the three paradigms differ systematically in the testing procedures. There are two main types of testing procedures: the *fixed-length trial* procedure and the *infant control* procedure. Familiarization studies use the fixed-length trial procedure, and habituation studies and violation of expectation studies use the infant control procedure. In familiarization studies, the infant is presented with a predetermined number of fixed-length trials. Each trial often starts with an attention-getter to fixate the infant's eye gaze toward the visual stimulus, and the visual stimulus continues to be present for the entirety of the trial length. Depending on the specific experimental criteria, the infants tested may or may not be considered as "habituated" after the pretest phase in the familiarization studies. During habituation and violation of expectations, people use the infant control procedure. This term refers to the fact that the infant's looking behaviors determine the length of each trial and the total number of the trials. Each trial ends when the infant looks away for some time (normally two consecutive seconds). When the trial ends, an attention-getter is automatically played until the infant fixates back to the visual stimulus. Once the infant looks back, the next trial will begin. The total number of trials that each infant will see also depends on his or her looking behaviors. Often, the researchers prespecify a "habituation criterion". If the average infant's looking time for some consecutive trials drops below the habituation criterion, then the pretest phase ends, and the infant is considered as "habituated".

Presentation styles also differ among paradigms. There are mainly two types of presentation: *serial presentation*, in which the infant is only presented with one visual stimulus in one trial, and *paired presentation*, in which the infant is often introduced to a pair of stimuli side-by-side at the same time. The presentation can differ in the pretest and test phases. The same or different presentation styles can be used in the pretest phase or test phase. In familiarization and habituation, researchers have used both the serial presentation and paired presentation for the pretest phase and testing phase respectively whereas the violation of expectation paradigm has only used serial presentation in the two phases.

The differences in implementational details, the procedures, and the presentation styles used all contribute to the differences in data interpretation. Researchers often disagree on the implications of the looking time paradigms: what does the difference in looking time tell us exactly? Some researchers ascribe a *lean interpretation*, attributing the differences to differences in perceptual capacity. Some researchers will adopt a *rich interpretation*, arguing that the differences among looking time reveal the different levels of infant's conceptual understanding. As Aslin has pointed out, the looking time paradigms face the "many-to-one" problem: too many postulated hypotheses, only one measurement available (Aslin, 2007). Oftentimes, it was not clear whether the tension comes from the paradigm itself or different views on the construct being measured.

For example, one venue for debate is the notion of "novelty". In test trials, infants' longer looking time is often associated with "novelty preference". But what is a principled way to define novelty. Consider two infants in a looking-time paradigm study. In the pretest trial, the two infants are presented with a stimulus S over and over again. Then, during the test phase, they are presented with stimulus X and stimulus Y, and they both look longer at the stimulus X than the stimulus Y. Assuming that both are showing a novelty preference by looking longer at X because X is more novel than Y when compared with S. However, the question remains because the visual stimulus X can be “novel” in multiple ways: is it novel perceptually, or is it novel conceptually? If looking time is the only measurement we gather, how can we tell which “novelty” or “familiarity” is the basis of the differences in looking time?

To address the issues in the interpretation of the measurements, many different theories and models on infants’ looking behaviors have been proposed. However, relatively few attempts were made to investigate whether the different theories were coherent. This theoretical vagueness directly leads to practical issues: researchers would often alter some parameters of the experimental procedure to adapt to the research questions, without questioning whether such adaptation would have a consequence on the meaning of looking time data. Therefore, the goal of this paper is to articulate the different theoretical assumptions typically associated with these three paradigms. I will start with a brief overview of different theoretical underpinnings behind the looking time paradigms. I will evaluate the explanatory scope, and then highlight the strength and weakness of each theory. Then, I will try to offer some suggestions for future researchers considering using visual paradigms.

2. Theoretical foundations Overview

Multiple theories and models have been proposed to account for the observed behaviors in infant looking time paradigms. The four most widely discussed ones are the Comparator Theory (Sokolov, 1963), the Multi-Factor Model (Hunter & Ames, 1988), the Object File Theory (the connection to the paradigms was made explicit in Baillargeon, Li, Luo & Wang, 2006) and the Dual Process Theory (Groves & Thompson, 1970). These theories attempt to connect the measurement of observed behaviors (i.e., how does the duration of looking time change or remain unchanged), to the interpretation researchers are endorsing (i.e., certain perceptual and conceptual capacities are or are not in place at a certain age). One should note that these theories are not, for the most part, mutually exclusive. This can be shown by using Marr's three-level explanatory framework. For any given cognitive phenomenon, we can analyze them in terms of the computational level, the representational and algorithmic level, and the implementational level (Marr, 1982). In analyzing the infants' looking behaviors, these proposed theories are mostly concerned with analyses on the representation and algorithm level. Each of them adopts different representations and different algorithms to account for the behaviors observed during a looking time paradigm. But what are the representations and the algorithms? How do different choices of the representations and algorithms lead to different theories and interpretations? Despite the popularity of infant looking time paradigms and extensive discussion of each of the theoretical models, surprisingly few attempts were made to connect these models altogether. Researchers rarely provide a theoretical justification of the paradigm choice (for a more detailed criticism, see Rubio-Fernandez, 2019). In this section, I plan to give a brief overview of each account and then highlight the connections (or disconnections) between these theories.

Regardless of which theoretical framework we are using, the theory should successfully explain the following phenomena during infants looking time paradigms.

1. Pretest: When shown repeated stimuli over and over again, infants' looking time during each presentation of the stimulus tends to decrease.

2. Test: Once the infant's looking pattern satisfies the criterion, when the infant is presented with two sets of the testing stimuli, the infant should in principle show one of the three reactions:

   a. Novelty Preference: the infant looks longer at the novel stimuli.

   b. Familiarity Preference: the infant looks longer at the familiar stimuli.

   c. Null Preference: the infant looks equally long at both the novel and familiar stimuli.

2.1 The Comparator Theory (CT)

2.1.1 Original Interpretation

Most authors in the literature attribute the first conceptualization of Comparator Theory to Sokolov's works on orienting reflex that date back to the 1960s. In the initial theory, Sokolov's conception was not specific to the human visual system. Instead, he argued that the orienting reflex is governed by an "independently functional" system in many organisms and its reaction "is specifically evoked by the novelty of the stimulus" (Sokolov, 1963). He also introduced the idea of "information value" into this theory. He believed that as the organism is repeatedly stimulated, it would form a model-like representation of the stimuli. When an incoming stimulus is similar to the built representation, it is considered as having less information value and thus leads to a decreased orienting response. But when the incoming stimulus is not similar to the built representation, it is considered as having more information value and thus leads to an increased orienting response instead (Sokolov, 1963).

In the original conception, orienting is characterized as a reflexive oculomotor behavior and an information processing mechanism at the same time. If one "orients" toward the stimulus, the subject must exert executive oculomotor control to turn one's eyes toward the stimulus, and as soon as the eyes are turned, the visual information would be processed into constructing a mental representation of the stimulus. When taken at face value, the original version of the Comparator Theory explained the phenomenon as follows: the decreased looking time during the pretest phase is due to a decreased orienting response. At some moments during the pretest phase, the infant no longer turns to the stimuli and no longer processes any information. Then during the test phase, when there is something new, the infant turns toward the stimuli again and starts to process new information, demonstrated by the infant showing a novelty preference.

The explication of this idea leads later research to focus more on infants' looking behaviors. Later research shows that the infants' looking behaviors are not driven by a single orienting mechanism. Research by Cohen (1972) shows that sustained attention also plays an important role in determining one's looking behaviors. He proposed that infants' looking behavior could be captured by two independent processes: an *attention-getting* process and an *attention holding* process. The attention-getting process resembles the orienting reflex in Sokolov's definition: infants turn their eyes toward the visual stimuli. But the attention-holding process is what determines how long the infant would fixate on the visual stimulus. Put differently, the total duration of looking time is jointly determined by a mechanism that orients the infant toward the stimuli and a separate mechanism that keeps the infant looking at the stimuli. This latter mechanism is the most relevant to the measurement of fixation duration.

In his study, Cohen found that the decrease in responses does not occur in the orienting phase of looking, but only occur in the sustained phase. He measured the latency of infants' orienting toward the stimuli and the duration of infants' looking time during the stimuli presentation. For the thirty-six four-months-old infants tested, Cohen found little correlation between the latencies and looking time duration. Later on, he and his colleagues confirmed that the decreased response toward stimuli was found in the attention-holding process (Cohen, DeLoache & Rissman, 1975). They discovered that with repeated presentations of stimuli, four-months-old infants only showed a decrease in their total durations of looking time, which is driven by the attention-holding process. The decrease is not found in their orienting latencies, which is driven by the attention-getting process.

Cohen and his colleagues' work replaced a simple and direct connection between orienting reflex and looking time with a further specified attention mechanism. The looking behavior became more widely accepted as a motor correlate of attention, and researchers made more progress in understanding the underlying mechanism of infant visual attention by introducing other measurements. One notable behavioral measurement is heart rate, introduced by the heart-rate defined multiphasic infant visual attention framework (Colombo, Richman, Shaddy, Follmer Greenhoot & Maikranz, 2001; Richards & Casey, 1991; for historical background, see Graham & Clifon 1966; Kagan & Lewis, 1965). The heart-rate defined attention phases divided attention based on the heart rate's change relative to the baseline heart rate. An infant enters the state of Sustained attention (SA) if he or she is looking, and during looking, his or her five consecutive heartbeats are below the baseline heart rate. The orienting phase (OR) and attention-termination (AT) are defined relative to SA. OR is defined as the period when the infants are looking at the stimulus, but the heart rate has not dropped below the baseline yet. AT is defined as the period when the infants are still looking but the heart rate has raised back to the baseline level.

Colombo and colleagues connected the more-refined framework of attention with the familiarization paradigm (Colombo et al., 2001). During the pretest phase, infants were presented with a one dot-matrix stimulus. The phase ended when the infants looked at the stimuli for 20 seconds accumulatively. After the pretest phase and a 650ms inter-phase interval, infants were presented with the pre-test phase matrix side by side with a new stimulus for either 5s or 15s of accumulated looking time. Infants' performance was reduced to a dichotomous variable: recognized or did not recognize. The “recognized” group were the infants who looked at the new stimuli for more than 55% of time, and this was taken as a piece of evidence for “novelty preference” because they showed an increased interest in new stimuli. The “did not recognize” group consisted of the remaining infants. Colombo and colleagues’ findings connected the attention phases and task performances in an interesting way. They found that only the duration of AT negatively related to the chance of recognition. In other words, the time spent in the orienting phase and sustained attention phase did not contribute to infants’ recognition performance. Only those infants who had longer time spent in AT were less likely to look longer at the novelty stimuli. This finding suggests the possibility that a third mechanism responsible for attention termination might be more closely related to the formation of representations during looking time paradigms.

The three accounts reviewed so far mark progress in their increasing partitioning of looking behavior, which in turn establishes a more fitting correspondence between motor behaviors and the underlying psychological capacity, attention. The framework developed by Colombo and colleagues aligns well with Cohen’s dual-process theory. It also adds to Cohen’s dual processes theory by revealing that the disengagement of attention is also crucial in determining the looking duration. Together, these results extend on the original conception of the Comparator Theory by specifying the components of underlying details. The characteristic behaviors observed in using the looking time paradigms (familiarization, habituation, and Violation of Expectation) are not fully captured by Sokolov's original focuses on orienting responses. The changes in looking time in the pretest phase and test phase were not simply due to the changes in orienting response.

2.1.2 Contemporary Interpretation

In the more contemporary interpretation, CT becomes more abstract. The emerging popularity of the information processing perspective revived and enriched the Comparator Theory. Now the infant's looking behavior is understood as an index of an encoding process, and the length of the encoding process is a function of the discrepancy between the encoded representation and the stimuli to be encoded (Bornstein, 1985). During the pretest phase, the infant looks less and less at the stimulus as the infant attains a better representation of the visual stimulus. In the testing phase, if the discrepancy between test stimuli and built representation is large enough, it will elicit a novelty preference and causes the infant to look longer at the new stimuli. In contrast, if the infant does not obtain a complete representation of the stimulus, the infant would continue to look at the familiar stimulus during the test phase and causes a familiarity preference.

However, this notion of "incomplete representation" is problematic. Comparator theory does not articulate an exact criterion for completeness. It assumes that the mental representations infants build have different degrees of "completeness", such that when the representation is considered as "complete", the infant would prefer to encode the new stimulus rather than spend more time elaborating on the representation of the current stimulus. But what does it really mean to have a “complete” representation of a stimulus? Is there a one-fits-all criterion for the “completeness” across the board? Or does the “completeness” of the representation depend on some specific features of the stimulus? Moreover, how do the infants know when the representation is complete and that “the habituation process has reached completion”? (Kavšek, 2013) Is there some pre-determined, built-in parameter that is embedded in each infant’s “comparator system”? These questions are important, but unfortunately they have been left unanswered in the comparator theory.

Despite the conceptual equivocation, the Comparator Theory still makes some progress with its more modern interpretations. Compared with the original conceptualization of CT, this new interpretation of CT treats attention as a more cognitive and internal process. Attention is no longer evaluated based on its motor correlates: the looking behaviors. In the more modern interpretation, when an infant is looking at the scene, one is actively processing and encoding the scene into one’s memory system. This newer version of Comparator Theory operates on two assumptions: first, when infants deploy their visual attention to the stimuli, they do not merely turn their eyes toward the stimuli. They attend to the stimuli. They actively analyze and selectively encode the relevant features of the stimuli. Second, the mental representation, which is the product of the encoding process, can be stored temporarily for future comparison. Both assumptions are supported by empirical evidence.

The active-processing view of infant visual attention is supported by studies in early categorization ability. Infants as young as three-months-old are capable of forming categories of cats and dogs based on perceptual features (Quinn, Eimas & Rosenkrantz,1993). In this study, the young infants were presented with different exemplars of pictures of cats or dogs in a side by side manner. Researchers found that after being familiarized with different exemplars of cats or dogs, these infants would show a novelty preference for pictures of animals from other species but not for the animals from the same species as the ones in the pretest phase. Note that if the infants were only encoding the features of individual exemplars of the animals, then they should show the same novelty preference toward both a novel exemplar from the same category and a novel exemplar from a novel category, because in both cases the perceptual features presented are different from the one that the infants were habituated to. However, the fact that the infants were only showing novelty preference toward a member of the novel category suggests that when the infant deploys their visual attention on the visual stimuli, he or she is not merely receiving the information passively. Instead, they are actively processing and extracting to encode relevant information for categorization when they are fixating on the stimulus.

Studies on infant memory also support the second assumption that even very early on the infants can maintain a relatively stable representation of the product of the processing. Researchers working on infant visual recognition memory have adopted the novelty preference and familiarity preference as “the defining features” of infant visual recognition memory (Rose, Feldman, & Jankowski, 2004). Most of the studies in infant visual memory involve some types of delayed response. For example, the classic “A-no-B” reaching task is widely used to study infants' memory performance. In this task, the experimenter shows the infant two pieces of cloth. First, the infant is shown an object being hidden underneath one piece of cloth repeatedly. The experimenter would encourage the infant to reach to the cloth and reveal the object. Then, the experiment will hide the object, in front of the infant’s eyes, under the other piece of cloth. Even if the infant sees the object being hidden in a new location, he or she reaches out to the old location. Previously, this “A-not-B” error led researchers to speculate that it was caused by an underdeveloped memory system, that the infant simply forgot where the object was. However, later research shows that even though the infants make persistent reaching mistakes, they are more likely to direct their gaze toward the correct location.

In a study conducted by Hofstadter and Reznick (1996), infants were shown repeatedly one toy being hidden in one of the two locations. Then, depending on the conditions, infants were encouraged to either reveal the hidden toy by pulling a cloth or search for the toy visually. Infants’ reach and gaze directions in these two conditions were coded for comparison. Interestingly, researchers found a discrepancy between the reaching and gazing conditions. Compared with infants in the reaching condition, Infants in the gazing condition were more likely to direct their eyes toward the correct location. This is considered as evidence for the infant’s ability to remember the location of the hidden object when it is out of sight, and their eye gazes can be considered as a reliable index for memory. The failure in reaching can be accounted for by other performance factors, such as timing, posture and prior experience (Smith, Thelen, Titzer, & Melin, 1999). Therefore, researchers agree that even very young infants can store, retrieve and act upon the products of the processing visual information.

In conclusion, as an analysis of the algorithm and representational level, the contemporary CT succeeds in explaining infants’ performance in the exposure phase and testing phase. It also receives empirical supports from studies on infant categorization and memory tasks. Nevertheless, CT still has one unaddressed weakness: the conceptual ambiguity of “complete/incomplete representations”. Therefore, it remains unclear what the criterion is for the comparator theory to determine when an infant should show novelty preference, familiarity preference or null preference.

2.2 The Multifactor Model

Compared with the Comparator Theory's decades of history, the Multifactor is young. It was first summarized in a seminal paper by Hunter and Ames in 1988. In this paper, the authors proposed that three factors jointly influenced infants' looking time behavior: age, task complexity and familiarization time (Hunter & Ames, 1988). Despite the brief history, this model has inspired a generation of infant researchers to be more careful with the experimental designs and data interpretation. Although this model was not presented quantitatively, it still became pivotal for researchers who aim to build a computational model for the infants' looking time behavior. When later researchers are trying to identify the criteria for the computational model for infants' looking time paradigms, they refer back to the characteristics identified in the Multifactor Model (Sirois & Mareschal, 2002, 2004).

The Multifactor Model tackles a specific subset of the behaviors in the paradigms: the factors influencing infants' behaviors during the test phase, either showing novelty preference, familiarity preference, or null preference. In the early discussion, I pointed out that any theory aiming to account for infants' looking behaviors during the looking time paradigms needs to account for the observed behaviors in pretest and test phase: the decreasing looking time during the pretest phase, and the different types of preferences during the testing phase. In this respect, it is fair to say that the Multifactor Model does not count as a full-blown theory in itself because it does not explain the behaviors in the pretest phase. Nevertheless, this limitation does not dismiss the value of this model. The Multifactor model still retains theoretical significance because it articulates factors that account for the relationships between three different preference patterns in the test phase. In this section, I will first briefly summarize the model and then discuss the model in the context of its application in the infant mathematical abilities.

2.2.1 ***The Factors in the model***

The first factor is the familiarization time. Hunter and Ames suggested the infants' preferences vary as a function of the duration of familiarization time in the pretest phase: namely, how long the infants are exposed to the stimuli. Researchers have long observed that in the looking time paradigms infants would show familiarity preference when they are not given sufficient familiarization time and show novelty preference when the familiarization time is extended (Wetherford & Cohen, 1973; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982). Based on these observations, Hunter and Ames established a preference curve (see Figure 2.a). Each infant starts with a null preference: if an infant is not given any familiarization time, then one is not going to pay selective attention to either “familiar stimuli” or “novel stimuli”, because for the infant the two of them are equally novel. As the familiarization time increases, the infant is going to look longer at the one she or he is familiar with. This is the phase of familiarity preference. Then, when the familiarization time further increases, the familiarity preference wanes, and the infant is going to have a second period of null preference. After this point, the infant is going to pay more attention to the novel stimuli.

Note that the comparator theory also implicitly suggests this shift from familiarity preference to novelty preference. A comparator theorist can state that when the infant has not yet formed a “complete” representation of the stimulus, she or he will keep looking at the stimulus and thus show a familiarity preference. When the representation is “good enough”, the infant will begin to look at the stimulus inconsistent with the built representation and thus show a novelty preference. However, as discussed earlier, the Comparator Theory has a very undefined notion of the “completeness” and “incompleteness” of the representation. In contrast, while the multifactor model does not explain the underlying mechanism for the preference shift, it does explicitly provide an operationalized criterion to capture such a shift, which is the duration of the familiarization time. The longer the familiarization time is, the less novel the stimulus is.

The second factor is age. Age is of special importance and relevance to developmental psychologists. As Hunter and Ames wrote at the beginning of the paper, one challenge for many developmental psychologists is to distinguish aspects of behaviors that “change with age from those that are age-invariant” (Hunter & Ames, 1988). Therefore, the differences between older infants and younger infants are important for the interpretation of the data. Hunter and Ames suggested that as the infants get older, they possess a richer knowledge base and a more efficient information processing style. In looking time paradigms, these two changes are demonstrated in the more rapid shift from familiarity preference to novelty preference in older infants. For example, with the same amount of increased familiarization time, a younger infant may still show a familiarity preference whereas the older infant has already entered the novelty preference phase (see Figure 2b).

Third, the nature of the tasks: the difficulty of a task in the looking time paradigms influences the direction of the infants’ preferences as well. Task difficulties in looking time paradigms come from three sources: the pretest phase, the test phase and the difference between the tasks used in the two phases. During the pretest phase, infants can be influenced by the design of the study. Some examples are how many toys the infants are exposed to, whether the stimuli are 2D shown on the screen or 3D real-world display and whether the stimuli are moving or remain still. These variations contribute to the complexity of the tasks for infants. During the test phase, the testing procedure differences also matter. For example, as Hunter and Ames point out, the serial presentation can be more difficult than the paired presentation (Hunter & Ames, 1988). In the serial presentation, the infants are expected to compare “a physically present stimulus with a mental representation of a stimulus”, whereas, in the paired presentation, infants are comparing two stimuli side by side. Finally, the differences between the pretest phase and the test phase also contribute to the task difficulty. An infant tested in an experiment that has similar procedures during the pretest phase and test phase shows a novelty preference more quickly than an infant tested in experiments using two different procedures in two phases. This is because of the contrast of the procedures themselves will impose more processing demands on the infant. Of course, it is noteworthy that task difficulty is age-relevant. The same task considered to be “difficult” for 3-month-olds can be “simple” for 8-month-olds. In the preference curve, the influence of task difficulties is reflected as the total duration of the sequence. When the age is held constant, the infants given a more difficult task would require a longer time to finish the shift from null preference to familiarity preference, and then to novelty preference (Figure 2c).

2.2.2 ***Application of the multifactor model: a case study in infants' arithmetic ability***

Literature in infant mathematical capacity demonstrates the theoretical importance of the Multifactor Model. In a controversial paper, Wynn (1992) claimed to have shown that five-month-old infants have the mental capacities to perform mental addition and subtraction by using the Violation of Expectation paradigm. In her studies, the infants were randomly assigned to two groups: the addition group or the subtraction group. In the addition group, the infants were seated in front of a stage with one object. During the test, a screen came up and occluded the object. Then, a hand holding an object entered behind the screen and left the screen empty. In the subtraction group, the infants were seated in front of a stage with two objects. During the test, a screen came up and occluded the two objects. An empty hand then entered behind the screen and left the screen holding one object. In both groups, the action sequences showed arithmetic operations. When each action sequence was finished, the screen was lowered to reveal the objects behind the screen. There were either one object on the stage or two objects on the stage. For the addition group, Wynn called the one-object result "impossible outcome" (1 + 1 ¹ =1) and the two-object result "possible outcome" (1 + 1 = 2). For the subtraction group, the "possible outcome" and the "impossible outcome" were reversed. The action sequence and result-showing phase were repeated six times, with three "possible outcome" trials and three "impossible" outcome trials. Infants' looking time during the test trials were measured and compared.

The results were interesting. Wynn found that these five-months-old infants looked longer at the "impossible outcome" than the "possible outcome". The infants in the addition group would look longer at the one-object trials than the two-objects trials, whereas the infants in the subtraction group would look longer at the two-objects trials than the one-object trials. Wynn took this as evidence suggesting that infants are performing mental addition and subtraction. These young infants calculated the number of the objects on the stage, and they would look longer at the outcomes when the outcomes were inconsistent with their expectations as if they were "surprised".

On the one hand, Wynn's study has been replicated multiple times and a meta-analysis has also shown that her original findings were replicable and reliable. (Simon, Hespos & Rochat, 1995; Christodoulou, Lac & Moore, 2017). On the other hand, the underlying mechanism of the looking time differences requires scrutinization. Wynn never justified why she believed that one set of outcomes was "novel" for infants. She tagged it as "impossible outcome", but it is only "impossible" for adults who already know math. To say that one set of outcomes was "impossible" for infants was to say that infants have already performed math. But this is what the experiment was trying to show. Therefore, we can only say that there were two different outcomes, and infants looked longer at one than the other. Now to interpret the difference, we must consider many different potential reasons. The two outcomes differed in many dimensions, both perceptually and conceptually. To say one is more "novel" to the infants than the other was to select only one specific dimension that the two outcomes differed. But Wynn never answered what justifies such selection.

Under the framework of the Multifactor Model, infants' longer looking time might be a result of perceptual familiarity preference. Some researchers have pointed out that in Wynn's study, the infants were young and were given relatively short familiarization time. Therefore, it is very plausible that they were driven by a preference for familiarity to look at the so-called "impossible outcome" (Clearfield & Mix,1999; Haith, 1998). In the addition group, the infants started with one object on the stage, and the so-called "impossible outcome" has one object on the stage as well. In the subtraction group, the infants started with two and the so-called "impossible outcome" has two objects as well. Therefore, the "impossible outcome" can be interpreted as perceptually familiar to the infants. They looked longer in one trial not because one plus one is not equal to two, but because one is one.

A series of experiments conducted by Cohen and Marks elegantly illustrated the evidence for a familiarity-driven account (Cohen & Marks, 2002). They considered three possible hypotheses: the familiarity driven hypothesis, the computation-driven hypothesis, and the directional-driven hypothesis. The familiarity driven hypothesis states that longer looking time is a result of perceptual familiarity with the stimuli; the computation-driven hypothesis, which is the perspective supported by Wynn, holds that infants look longer at certain scene when the scene is not consistent with their results of mental computation after seeing the action sequences; the directional-driven hypothesis is a weaker version of the computation hypothesis, which states that infants know that addition leads to more and subtraction leads to less. When the direction of change does not fit their expectations, they would look longer at the scene. To examine the three hypotheses, Cohen and Marks designed four different arrays that the infants see after the action sequences: no object, one object, two objects or three objects.

Each hypothesis leads to different predictions about infants' looking time during the four different results. If infants are driven by familiarity preference, they would only look longer at the outcome perceptually similar to the one they see before the action sequence started. The other three outcomes should be equally novel and create equally short looking time. If infants are driven by their computation, then they should look equally long on all the wrong results but look selectively short on the correct result. Lastly, if the infants are driven by the directions of the change they detected, then they should look equally long at the no change or the wrong direction of change conditions. For example, in the addition condition, the familiarity-driven hypothesis predicts that the infants would look longer at the one-object trials while equally short at the zero, two and three-objects trials. The computation-driven hypothesis predicts that the infants would look longer at zero, one and three-objects trials, compared to the two-objects trials. The directional-driven hypothesis predicts that the infants would look longer at the zero and one-object trials than the two- and three-objects trials because the direction of change is inconsistent with what the infants expect.

The first and the third experiments from the paper addressed the possible hypotheses. In the first experiment, Cohen and Marks implemented the studies as similar as possible to the original Wynn's study. The only difference is the number of possible outcomes that young infants would see (now four instead of two). In the third experiment, Cohen and Marks omitted the action sequence. Infants were directly presented with either one object on the stage or two objects on the stage. Then, a screen rotated up and occluded the object. When the screen rotated down again, the infants would see one of the four possible outcomes. Intriguingly, in both experiments, Cohen and Marks found that the infants would look longer at the one object when they started with one object, regardless of whether there is an action sequence or not. They conducted a post hoc statistical analysis by selecting 16 infants from the first experiment and compared their results with the infants in the third experiment. The patterns of looking across the two experiments were the same: they would look longer at the familiar scene than the unfamiliar scene.

Cohen and Marks concluded with an application of Occam's Razor: "When certain abilities are attributed to young infants, simpler mechanisms can sometimes account for the data" (Cohen & Marks, 2002). They took the findings in their experiments as showing that a perceptually based familiarity preference account is sufficient to explain the behaviors observed in Wynn's experiments. This suggestion shows that even though the multifactor model itself does not offer a mechanistic explanation, it does motivate researchers to consider a simpler explanation. When a simpler, perceptual explanation is available, one should not seek the more complex, conceptual explanation. Our understanding of the infants' conceptual life is based on inferences from observed behaviors. So, it was important to justify and constrain our inference on the ground of evidence.

The debate around Wynn's findings was only one manifestation of the many positive influences that the Multifactor Model brings. This model provides a systematic way to evaluate looking time based on the three factors: age, familiarization time and task difficulty. Developmental psychologists are interested in developmental change, so age has been the major factor under consideration. Yet experimental methods also influence infants' behaviors in significant ways. Therefore, to interpret data gathered from looking time paradigms one has to consider how differences in familiarization time and task difficulty would contribute to the observed infants' behaviors. Just as Oakes and Mandole pointed out, researchers should interpret their looking time data in a "procedure-oriented" way instead of a "content-oriented" way (Oakes & Mandole, 2000). The content-oriented way refers to treating the looking time methods as a clear window to infants' knowledge about the world. This perspective is misleading because the methods themselves are procedures that infants can learn through and be shaped by. In conclusion, looking time data's interpretation requires consideration of the influence of experimental procedures.

2.2.3 *Implications of the Multifactor Model*

The Multifactor Model is constructive as a framework for interpreting looking time data. But it does not offer any mechanistic explanation as to why and how the behaviors observed emerge. This model only becomes more explanatory when combined with mechanistic theory such as the Comparator Theory. Nevertheless, the Multifactor Model adds three important parameters when considering actual mechanistic explanations. In the following, I will use the Comparator Theory as an example, although note that the Multifactor Model is a flexible “add-on” to other mechanistic accounts of the looking behaviors as well.

First, the total duration of looking time ties into the preference pattern in the test phase. While the comparator theory itself only has a vague account based on “the completeness of the representation”, the Multifactor Model operationalizes it into the total duration of the looking time. The second factor, age, is connected to how fast the looking time would drop. The Comparator Theory does not articulate how the comparison process would alter as the infants grow, but the Multifactor Model articulates that age is related to the steepness of the preference-shift curve. The older the infants are, the faster the looking time would decrease. The third factor, task complexity also remains unmentioned in the Comparator Theory, but the wide varieties of test procedures, task content and implementational details in infant looking time paradigms require special considerations in data interpretation. In this respect, Multifactor Model leaves a cautious note on the influence of task difficulty: when the age is held constant, the more difficult tasks require a longer time to accommodate the decreased looking time and the familiarity-to-novelty shift.

So how does the coupling of the two theories inform us about the three paradigms and their application? First of all, task difficulty should be taken into considerations when researchers are designing experiments, especially for studies using the Violation of Expectation paradigm. As reviewed earlier, the debates on infants' mathematical abilities are partly motivated by the considerations based on task difficulty. Wynn's experiments, common in many studies using Violation of Expectation Paradigm, involve 3D dynamic events and serial presentations. These implementations cause the task to be more cognitively demanding than 2D static display and side-by-side presentations more commonly seen in habituation and familiarization paradigms. As a result, researchers should make sure that the infants have received an adequate amount of familiarization time to make sure that they have shifted from the familiarity preference phase to the novelty preference phase. This is often achieved by either extending the familiarization time or using the more stringent criteria for the infants' looking time declination.

Second, the age factor also sheds insights into the evaluating of the validity of the measurement yielded by the looking time paradigms. Age can influence the results in two ways: the different task difficulty and the different mechanisms being studied. As discussed earlier, the task difficulty itself is an age-relevant factor. While the same task can be difficult for younger infants, it might be considered as relatively simple for older infants. However, few attempts are made to address the variation of task difficulty for different age groups. The one experimental study that investigated this issue experimentally only has a very rudimentary notion of simple and complex: in one of the earlier works by Hunter and colleagues, they adjust the number of toys to construct "difficult" and "simple" tasks. They consider a three-toy-array as "difficult" for eight-month-olds, and "simple" for 12-month-olds while the "difficult" task for 12-month-olds is a "five-toy-array" (Hunter, Ames, & Koopman, 1983). Without doubt, more work needs to be done before the field can rely on a more precise, operationalized concept of age-relevant task difficulty.

The other age-related concern is the different mechanisms being tested in one study. The assumption behind many looking time paradigms is that the looking time measurement T is the behavioral correlate of one cognitive or perceptual mechanism M. The differences among results collected, R1 and R2, are the reflection of the different levels of maturity of this mechanism M. However, this assumption can be problematic. In some cases, even in the same experiment, infants of different ages can rely on different mechanisms either due to different task difficulties or due to the maturity of a separate mechanism. Infants of different ages may use two different mechanisms, M1 and M2, in the same task. For example, Carey and Xu (2001) argued that the different performance in infants' object individuation task can be understood as due to two different mechanisms in operation in different age groups. According to them, infants younger than 10-month-olds may rely on the mid-level object file system to support their object representations, whereas infants around 12-month-olds develop a second object individuation system: the kind-base system. Thus, the difference between performance across two age groups need not be due to the differences in the level of maturity for one system. Rather, it can be the result of different mechanisms are used. This possibility, again, emphasizes the importance of taking into consideration age factors when using the looking time paradigms.

While the insights brought by the Multifactor Model are valuable in many ways, some weaknesses within the model require scrutiny. To start with, while the notion of novelty as equivalent to the exposure time has the virtues of being intuitive and operationalizable, it also has an unjustified assumption: the infants' personal experience before being brought into the lab will not interfere with their "starting points" with certain stimuli. This assumption does not always stand, however. For example, if an infant comes from a household with a dog, she or he might be already familiar with the visual features of a dog and thus consider a picture of a dog as less novel than a picture of a cat. In contrast, an infant who lives in a household with a cat may find a picture of a dog as more novel than a picture of a cat, and therefore during the experiment, he or she will need more time to make the dog as equally "familiar" as the former infants. Admittedly, the experimental method in psychology can hardly rule out the influence of personal history outside the lab. It can only be mediated by randomized sampling and control groups for comparisons. However, as researchers studying infants are in search of the "default state" of human cognition, the implicit assumption that "all infants are equal" in relation to the novelty and familiarity of specific stimuli challenges the potential generalizability of the findings because this assumption can only stand when the experiment's sample is highly homogeneous in background.

Another limitation that the Multifactor Model suffers from is the lack of mechanistic explanations it provides for the looking time behaviors. As Carey pointed out in her comments on the Cohen and Marks paper, the familiarity-driven account based on the Multifactor Model still needs to address the underlying mechanism. Cohen and Marks' interpretations were missing the necessary details: they never specified "the format of representation", "the computations involved" and "the nature of the comparison" (Carey, 2002). This criticism is hard to dismiss because there is no justification for why the Multifactor Model has to commit to the Comparator Theory as *the account* for the underlying mechanism. On the one hand, the multifactor model is able to accommodate different theories because it emphasizes correlates rather than the explanations of the underlying mechanisms. On the other hand, this flexibility has a dark side, which is its conceptual vagueness that causes the arguments built upon to remain unconvincing.

In summary, Multifactor Model identifies three important factors to consider when interpreting the looking time data: familiarization time, age and task difficulty. For paradigms that involve more complicated settings and more difficult tasks, such as Violation of Expectation, one needs to make sure that the participant has received adequate familiarization time for the stimuli presented. What's more, the age can also interact with the task difficulty, and might even alter the mechanisms the task tapped into. Nevertheless, the connection between age and task difficulty remains speculative and relatively vague. Another ambiguity is in using familiarization time as a proxy for the familiarity and novelty. Such connection is questionable because it assumes that all infants, regardless of their prior experience, start from the same level of novelty for any stimuli being shown. This assumption can limit the generalizability of the findings because it requires the participant sample to be highly homogeneous in their background. In conclusion, More work needs to be done before researchers can account for and depend on the systematic relationships connecting the age and the task difficulty, as well as the familiarization time and "novelty".

2.3 Object File Theory

To see how Object File Theory accounts for infant looking time paradigms, we need to first dive into the notion of Object File. This concept originated from studies on adult object-based attention. In a seminal paper by Kahneman and colleagues., they presented evidence suggesting "object-specific perceptual representations", which were named "Object Files" (Kahneman, Treisman & Gibbs, 1992). In one study, subjects were seated in front of a screen. Eight squares were presented on the screen. During each trial, participants would see a few letters flashed up inside the square. Then, one single letter flashed up in one of the squares, and the subjects were asked to name the last letter as quickly as possible. Kahneman and colleagues found that the subjects were faster to respond when the letters appeared in the same square before, and thus showed an evidence for object-specific benefit. Such benefit was interpreted as indicating that the attention allocation process should be considered as object based, which gave rise to the notion of "object file". Although the initial definition for Object Files provided by Kahneman and colleagues was rather vague, i.e. "a temporary episodic representation", it still sparked decades of empirical research investigating the attentional process driving the operations of Object Files. In a more recent theoretical review, this issue was examined again. Basing on almost three decades of empirical research, Green and Quilty-Dunn proposed a new framework to understand Object File. They summarized that Object Files should be understood as a propositional representation that "i) sustains reference to an external object over time, and ii) stores and updates information concerning the properties of that objects." (Green & Quilty-Dunn, 2017).

Since physical objects are everywhere in our daily experience, then perhaps unsurprisingly, psychologists would be interested to see how infants represent them. The earlier works of Piaget suggested that object concepts emerged relatively late, not until infants reach around 2-years of age. (Piaget, 1954). But starting from the 1980s, more and more research, utilizing Violation of Expectation paradigm, began to provide evidence that infants during the first year of infancy are capable of representing objects (Baillaregeon, Spelke, and Wasserman, 1985; Bailargeon, 1987; Bailargeon & Graber, 1987; Bailargeon & Devos, 1991). However, given the ambiguous nature of the looking time data, the findings were often very controversial. On the one hand, people were arguing for the conceptual nature of the initial object representation. For example, Spelke argued that young infants "endow the world with entities that are *cohesive, bonded, substantial and spatiotemporally continuous*" (Spelke, 1988, italics from the original). Neither of these properties is available directly in the sensory input, but they constrain infants in their determination of the unity of objects. More importantly, this theoretical orientation grants infants to have beliefs about objects, similar to how adults having beliefs. (e.g., we may hold the belief that "Object X is solid so Object Y cannot pass through it.") In contrast, people against such a conceptual account would call for a deflationary explanation of the looking time differences, attributing the observed patterns to perceptual input. For example, Bogartz and colleagues questioned the empirical evidence that the conceptual argument is based on (Bogartz, Shinskey, & Speaker, 1997). According to them, many Violation of Expectation studies did not adequately control for the influence of perceptual novelty, and thus lead infants to look longer at some trials than others. They also presented some initial evidence suggesting that when the perceptual factors were controlled for, the evidence supporting the conceptual account vanished.

Object File Theory was proposed as the third possibility between the perceptual-conceptual dichotomy. This idea was inherited from the conception of visual attention as an interface between perception and cognition (Julesz, 1990; Pylyshyn, 1999). As units of object-based attention, Object Files are neither perceptual nor conceptual. They are not perceptual since they are stable, discrete representations. It is unclear that how they could come out of the fleeting, continuous sensory information. At the same time, they are not conceptual because they can represent objects without any identity information of the objects. In the playful words of

Kahenman et al., object files enabled us to exclaim "It's a bird; it's a plane; it's Superman!" when all the pronouns in these sentences refer to the same object. This is to suggest that we can individuate one entity as an object without identifying such an object as a particular kind. For some proponents of Object File Theory, infants' initial representation of the objects is in the format of these object files. The operations of these object files are also speculated to be the underlying mechanisms of infants' looking behaviors during the looking time paradigms.

In this section, I will first elaborate on the object file theory in the context of object-based attention as well as the infants' object concept. Then, I will compare this new theory with the Comparator Theory and Multifactor Model to show their connections and incompatibilities. I argue that for the most part, the Object File Theory is not incompatible with the previous two theories. However, while the Comparator theory partitions attention into a multi-phasic construct behaviorally and supplements it with cognitive processing, the Object File Theory is a more mechanistically specified theory that with a better-grounded theory of attention. Following that, I will highlight the major limitations of this theory: the more restricted explanatory scope. In the end, I will draw a cautionary conclusion about methods.

2.3.1 Object file theory: From Object-Based Attention to Object Concepts in Infancy

Attention is a ubiquitous construct. It contains a wide range of heterogonous perceptual and cognitive processes. Broadly construed, a taxonomy of attention has two major branches: external attention and internal attention (Chun, Golomb & Turk-Browne, 2011). The key distinction between these two branches of attention is the target of selection: the internal attention operates on the information generated internally, such as from memory. External attention operates on sensory information such as visual sensory input. Each category can be further partitioned into more specific categories. For internal attention, it can be partitioned by tasks, such as for rule switching or option selection. For external attention, it can be partitioned by modalities. Or it can also be partitioned by the selection unit: when the information is selected, what is the "unit" of selection? For visual attention, it can be based on space, time, feature and object.

Object File Theory was first proposed as a mechanism for object-based attention. In a seminal paper, Kahneman, Treisman, and Gibbs (1990) characterized an object file as "a temporary episodic representation, within which successive states of an object are linked and integrated". A different, but closely related theory is the FINST visual indexing theory (this abbreviation stands for FINgers of INSTantiation) proposed by Pylyshyn. (Pylyshyn, 1989; Pylyshyn, 1994). The FINST visual indexing theory holds that our visual attention works analogously like human fingers: it can point at objects in the environment, tracking them continuously through time and space without revealing the features or the identity of the object. The FINST visual indexing theory and the Object File Theory are often considered as different parts of one indexing system. Object Files are considered to contain more featural information about the objects whereas FINSTs have the spatiotemporal addresses to the objects being indexed.

Many have speculated that Object Files were the underlying representations of infants' initial object concepts. Carey and Xu (2001) discovered three similar features of infants' looking behaviors in object concept studies and adults' behaviors in object-based attention tasks: the primacy of spatiotemporal information, the set size limitation and the occlusion vs existence cessation. The first feature is that, similar to adults, infants seem to prioritize spatiotemporal information in determining the numerical identity of the objects. Xu and Carey (1996) found that infants at 10-month-olds can only judge the number of objects accurately when given strong spatiotemporal evidence. In one experiment, the infants saw a sequence of events. First, a duck moved from behind the screen, paused, then returned and disappeared behind the screen. Then, a ball repeated the same event sequence. When the screen was lifted, either two objects, a duck and a ball, or one object, only the duck, was revealed. Infants at 10-month-olds did not respond differently to the two conditions. This was taken to be the evidence suggesting that they did not encode the numerical information of the objects. They did not have an expectation of how many objects were there behind the screen even if they saw the two action sequence a moment ago. However, in another condition, their reactions changed when they saw the duck and the ball moved in the same action sequence simultaneously. Now, they would look longer at the scene in which only the duck was revealed compared to both the duck and the ball were presented. The difference between the two conditions is the presence of spatiotemporal information. When the two objects were involved in the action sequence at the same time, they provided strong spatiotemporal cues to the numerical identity. Therefore, the infants would discriminate between when either one or two objects were shown. In adult object-based attention literature, it has been well-established that spatiotemporal identity enjoyed a similar privileged status. For example, the classic “tunnel effect” refers to the phenomenon where adults report perceiving one object moving continuously when they see an object disappear behind a barrier on the one side and later emerge from the other side. Even if there is a disruption in the sensory information about that object, people still perceive it as being the one object. This is also known as “amodal completion”, and it was taken to be evidence for the primacy of spatiotemporal cues in determining visual objecthood (Burke, 1952).

A second parallel between infants’ object concept and the object file system is the limitation of the set size. People have long noticed that even very young infants have the ability to discriminate different numbers of items. Starkey and Cooper (1980) were among the first to show that infants about 5.5-month-olds can discriminate between two objects and three objects but failed to do so when presented with four objects and six objects. Object-based attention, on the other hand, has a similar restriction on the number of objects trackable. The Multiple Object Tracking (MOT) Paradigm pioneered by Pylyshyn provided ample evidence suggesting the limited set size is also around four. In a typical MOT experiment, adult participants were shown a few objects presented on the screen. The participants are told to track the “target” objects, which are the objects that flash at the beginning of each trial. After the flashing stops, all objects start to move randomly across the screen. When the motion stops, the participants are asked to identify the flashed targets from the beginning of each trial. Pylyshyn and Storms (1988) showed that adult participants can track about four targets at the same time. When there were five targets flashed, the percent of error increased significantly. The similar upper limit found in infants’ number perception and object file theories supports the view that the object file system supports the infants’ representation of objects.

The last venue for converging evidence proposed by Carey and Xu is the sustained existence of objects in occlusion conditions. When an object is occluded, does it still exist? According to Piaget, infants developed the ability known as “object permanence” relatively late. When an object is occluded, an infant would think it has ceased to exist. But this view had been overturned by Baillargeon, Spelke, and Wasserman (1985) in their famous drawbridge experiments. In this classic Violation of Expectations study, infants as young as five-month-olds were shown two types of event: the impossible event and the possible event. In the experiment, a block is presented and then occluded by a screen. In the impossible event, the screen fell onto the block and seemingly "passed through" it. In the possible event, the screen stopped in the middle of the fall, as if being "stuck" by the woodblock. According to Baillargeon and colleagues, if the infants understand that the block occluded by the screen still exists, they should be "surprised" by the impossible event because the movement of the screen violates a basic property of the physical object. And indeed, the tested five-month-olds infants systematically looked longer at the impossible event than the possible event. Baillargeon and colleagues took this longer looking time to be evidence suggesting that infants' have a basic understanding of objects' physical properties. Of course, just like the perceptual-conceptual debates in infants' arithmetic experiment, this interpretation of longer looking time is also open to a lower, perceptual alternative explanation (Rivera, Wakeley, & Langer, 1999). However, Scholl and Leslie (1999) pointed out that there is a third option beyond the dichotomy between perception and cognition: the object concepts as object files. Evidence from the MOT paradigm has shown that these intermediate representations of objects can survive occlusion as long as the spatiotemporal cues to objecthood were provided. (Scholl & Pylyshyn, 1999). Therefore, infants' prolonged looking time toward the impossible event might be caused by the operations of the object files system.

2.3.2 Object File Theory's Relations to the Method and Other Theories

If infants' object concepts are realized by object files, and we study infants' object concepts by the looking time paradigms, then the Object File Theory must be able to accommodate the looking patterns in looking time paradigms as well. In other words, the object file theory should explain the infants' looking behaviors: why infants show the looking patterns in the pretest phase and the test phase. Furthermore, if we assume that the Object File theory can explain the observed behaviors, then to what degree is this theory compatible or incompatible with the other theories. How does the object file theory relate to the comparator theory? How does the Object File theory connect to the Multifactor Model?

Similar to other looking time experiments, object concept studies have two phases, a pretest phase and a test phase. Typically, during the pretest phase, the infants were introduced to the objects of interest. However, the looking time data during this pretest phase is treated and interpreted very differently across experiments and researchers. Some more conventional studies use a "habituation criterion" for pretest trials. In the classic experiment by Spelke and colleagues, infants' looking time during the pretest phase was collected and calculated to determine when the pretest phase ends and test phase starts (Spelke, Kestenbaum, Simons, & Wein, 1995). The pretest phase ends when a habituation criterion was met. In this case, it was a significant drop in looking time, such as the average of three consecutive trials' looking time is less than 50% of the first three consecutive trials. The pretest phase can also end when the maximum numbers of pretest trials had been administered, whichever came first. However, there are also works on infants' object concepts that discard the habituation criterion. Some researchers only focused on the looking time during the test phase, not even reporting on the looking time data during the pretest phase (Káldy & Leslie, 2003). Others treat the usage of the pretest phase as irrelevant to the test phase. For example, in Hespo and Baillargeon's study, they use only one single pretest phase, yet still were able to acquire consistently interpretable data from the experiments (Hespo

& Baillargeon, 2001). In their studies, four pretest trials were followed immediately by the six test trials, regardless of how long infants look in each trial as long as the infants go through the test phase. In a similar spirit, some researchers completely discarded the pretest phase. In a study testing infants' understanding of the hidden object, the researchers only presented each infant with two test trials and compared their looking time across the two trials. With the single exposure to the scene, the researchers still found infants looked significantly longer in the trial that contained the event that disobeyed the physical properties of the objects (Wang, Baillargeon & Brueckner, 2004).

The great flexibility in the procedure drew a lot of attention from the field, especially onto this latter, non-habituation version of Violation of Expectation paradigm. First of all, some may consider the looking time paradigm's two-phase structure as a definitional characteristic. An implicit assumption is behind all three paradigms: there is a causal connection between infants' behaviors in the pretest phase and the test phase. When this causal connection is set aside, it is questionable whether such a testing procedure still belongs to the same category of paradigm. Second, this approach to looking time paradigm shifts the conceived function of the testing procedure. As Oakes and Mandole (2000) pointed out, one can conceive looking time paradigms in two different ways: to adopt a content-oriented approach or to adopt a process-oriented approach. Content-oriented approach views looking time paradigms as a mean to shed light on the psychological capacities infants already possess before coming into the lab. In contrast, the process-oriented approach treats the paradigms as a process in which infants learn from. For the latter, the test phase of the paradigm tests on the content of "on-site learning": what the infant has learned through the pretest phase. Due the individual differences in learning rate among infants, the habituation criterion is necessary because it is a way to make sure that all infants have completed the learning process adequately. Only when all infants have reached the same habituation criterion does it make sense to compare the content of learning when all the infants being tested reached the same habituation criterion. To put it differently, the habituation criterion marks the standard of learning process for all infants. When the habituation criterion is discarded, there is no way to assess whether all infants have made the comparable process to support a reasonable comparison between the learned content. Therefore, if one uses looking time paradigms without a habituation criterion, one must be taking the content-oriented approach, which is discouraged by Oakes and Madole (2000).

It would be difficult to respond to the first issue without diving into a discussion on the definition of the method and the ontology of experimental paradigm. Such discussion is interesting in its own right, but it can provide minimal guidance to empirical research. For the second issue, a response could be that there is no intrinsic ordering of the two perspectives on looking time paradigms. All researchers must agree that infants are endowed with *some* psychological capacities prior to coming into the experiment. This is not to say that the infants possess a full-blown conceptual capacity to engage in physical or metaphysical reasoning. Nor is it to commit to a nativist position and regard all the psychological capacities shown in labs as being present at birth. For example, Object File Theory, as an alternative between the perception-cognition dichotomy, has no explicit commitment to its ontogeny and does provide a reasonable justification for the content-oriented approach. Few would argue that infants learn to see objects during the testing procedure.

The operation of object files can account for the infant looking behaviors. Again, take Spelke and Colleague's classic experiment as an example (Spelke et al., 1995). In this study, infants either saw a continuous event or a discontinuous event. The infants who saw the continuous event looked longer when they saw that there were two objects behind the screen. Meanwhile, infants who saw discontinuous event looked longer when they found only one object behind the screen. Object File theory would explain this looking pattern as follows: the increased looking time is caused by infants opening different numbers of object files when viewing different scenarios (Carey & Xu, 2001; Scholl & Leslie, 1999). In the continuous event, only one object file was opened during the pretest since the object files can sustain occlusion as long as the infants were given the appropriate spatiotemporal information (Scholl & Pylyshyn, 1999). When in the test phase two objects were revealed, the additional object triggered an additional object file to form. In other words, the longer looking time is the behavioral correlate of the formation of the new object file. Similarly, the discontinuous event in the pretest leads infants to open two object files. When only one object was shown in the test phase, infants increased their looking time to "search for the 'missing' object corresponding to the ongoing index" (Scholl & Leslie, 1999). In these cases, a habituation criterion during the pretest phase is no longer necessary. Object files often operate on the timescale of hundreds of milliseconds. There is evidence suggesting that it takes only up to 200 milliseconds for the visual system to compute a bounded objecthood from spatial relations (Feldman, 2007; Kahneman et al., 1992). Therefore, upon initial viewing of the scenes, the infants have established corresponding Object Files. The repeated presentations of the stimulus would be unnecessary.

Interestingly, this notion of "mismatch" in object file theory has a striking resemblance to the Comparator Theory. Kahneman and colleagues described the recognition process as following:

> "To mediate recognition, the sensory description in the object fields is compared to the stored representation of known objects. If and when a match is found, the identification of the object is entered in the file, together with information predicting other characteristics, its likely behavior, and the response it should appropriately evoke."
>
> Kahenman, Treisman & Gibbs (1992), p178

Compared to Sokolov's description of orienting reflex:

> "The orienting reflex as a complex functional system includes the integrative activities of different brain areas. Its distinguishing characteristic is that it arises in response to novelty. It depends upon elaboration of a nervous model of stimulus and the mismatch between the model and a new stimulus. The elaboration of the neuron model consists in fixation by the nervous system of stimulus traces. The origin of the orienting reflex apparently lies in a mismatch of extrapolatory impulses and afferent signals reaching common efferent neurons."
>
> Sokolov (1963), p576

Object File theory, similar to the Comparator Theory, holds that there is a comparison process underlying the change in the behavioral responses. The important difference between the two theories is the representation of what is being compared. The Comparator Theory vaguely evokes the "nervous model" or "representation", whereas the Object File Theory gives a more specific description of the nature of such representation. In a more recent article, object file's representation format is described as being propositional and contains multiple-slots for storing features of different categories (Green & Quilty-Dunn, 2017). This level of specificity is advantageous since it then provides an opportunity to spell out the detailed mechanism of the comparison. As discussed earlier, the weakness of Comparator Theory lies in its equivocation about what it means for a representation to be “complete” or “incomplete”. But Object File Theory does not have the same problem: the comparison is driven by the number of object files operating, as well as the features of the objects encoded in the object files.

Object File Theory is not in direct conflict with the Comparator Theory on the mechanistic level. The two important differences distinguishing these two theories are the timescale and the representations. The Object File theory’s differences also bleed into its compatibility with Multifactor Model, which ascribed three factors in interpreting the looking time results. These three factors are familiarization time, task difficulty and age. With Object File theory, familiarization time is no longer applicable since object files opened up instantaneously. Nevertheless, task difficulty and age still influence infants’ performance, determining whether or when certain features got encoded in object files. Baillargeon (2004) observed that infants of different age responded to an object’s spatiotemporal continuity differently. Infants at 4-month-olds would only consider the size and shape of the object. At 7.5-months they start to encode patterns of the objects in the object files. And eventually at 11-months color is also encoded in the object files (Bailargeon, Li, Luo & Wang, 2006). This series of developmental change suggest that for Object Files Theory task content can still impose different levels of task complexity on infants of different ages. For those who adopt Object File Theory, age and task difficulties are still two factors constraining the interpretations.

2.3.3 Limitation: Beyond Object Concepts and Physical Reasoning

All the previous experiments concern the infants’ understanding of the object concept. But this is certainly not the only place where infant looking time paradigm was applicable: from categorization ability to causal perception, from goal-directed action to theory of mind, the omnipresence of infants' looking time paradigm challenges the Object File Theory. To what degree can a working mechanism of visual attention accounts for all the other findings in infant cognition? Indeed, object files break through the dichotomy of perception and cognition in one area of infant research, but is there an analogous alternative for the debates on some more abstract abilities – the formation of category, the understanding of fairness, the attribution of false-belief – do all the findings from these areas have similar middle grounds between pure low-level perceptual explanation and high-level conceptual understanding? For some areas, maybe. For example, in the study on infants' category formation, the categories can be understood as built from the encoded features in the object files. But in this case, we are still testing infants' ability to process some perceivable features. What if the subjects of interests are unperceivable mental states?

One line of research pioneered by Woodward used the habituation paradigm in testing young infants' understanding of goal-directed action. In the seminal paper, infants of 5-month-olds and 9-month-olds saw two objects side by side. In one condition, during the pretest phase, the infants repeatedly saw a hand reaching toward one of the two different objects. The object being reached was considered as the "goal" object. This study adopted an infant control procedure. Each trial ended when the infant looked away for more than 2 second or 120s had lapsed. The pretest phase ended when the predetermined habituation criterion was met: the sum of three consecutive trials was less than the sum of the first three trials longer than 12 seconds. Once the pretest phase ended, the infants immediately entered the test phase and saw two types of trials: a "new goal/old path event" or an "old goal/new path event". The set-up of the test phase was identical to the pretest phase except for the two objects had been switched. The left toy was now on the right and vice versa. In the “new goal/old path event”, the infant saw the hand reaching toward the same location which was now occupied with a new object. In contrast, in the “old goal/new path event”, the infant saw the hand reaching toward a new location which was now occupied with the old object. Woodward found that infants as young as 5-month-olds looked significantly longer during the trials in which the hand was reaching in the same path toward a different object. Interestingly, when the reaching entity was a rod or a mechanical claw, the same pattern did not emerge and the infants look equally long at the two types of trials (Woodward, 1998).

This research has been replicated and extended across different paradigms in different labs (see Woodward, Sommerville, Gerson, Henderson, & Buresh, 2009, for a comprehensive review). Notably, in one replication attempt, Biro and Leslie (2007) used the familiarization paradigm instead of the habituation paradigm. They did not set a habituation criterion. In their studies, 12-month-old, 9-month-old and 6-month-old infants saw a stage similar to the one in Woodward’s. A hand or a rod repeatedly reached for one of the two different objects. All of the infants were shown four trials of reaching, each lasted about 20 seconds regardless of their own looking. Then, the test trials began, and the infants saw two types of trials, one goal-changed and the other path-changed. Despite the methodological differences, Biro and Leslie found a looking time pattern similar to Woodward’s. Infants in all three age groups looked longer on average at the goal-changed trials than the path-changed trials. They interpreted the evidence as suggesting that infants as young as 6-month-olds can “evaluate the goal-directedness of the observed actions”.

It is unclear how the Object File Theory could explain these converging findings. Researchers are finding similar patterns with two different procedures. One with the habituation criterion. The other without. Is it also a case where the Object File System is operating? Unlikely. There are no mismatches in the number of objects seen by the infants. Across the conditions and studies, two objects were repeatedly reached by another object (the hand). Presumably, these lead the infants to open three object files. During the test phase, however, the scene still consists of three objects. There is no mismatch in the number of object files and thus it is unlikely to be the cause of extended looking time. One may argue that the "goal" of the reaching hand might be encoded in the object files. It has been suggested that our perception of animacy and causality might be deeply grounded in our object-based attention system (Scholl & Tremoulet, 2000; Gao, Newman & Scholl, 2009; Newman, Choi, Wynn & Scholl, 2008). If properties as invisible as anomaly and causality can be grounded in our visual attention system, then it opens the possibility for "goal" to be encoded in the object file representation as well. Moreover, Leslie has proposed that infants are born with a Theory of Mind Mechanism (ToMM) that can "function relatively spontaneously since it has the job of directing the child's attention to mental states" (Leslie, 1995; Leslie, Friedman & German, 2004). Therefore, it is at least possible that when viewing the reaching scenes, some portion of infants' object-based attention is allocated to this unobservable "goal", and this "goal" enters the infants' representations of the event sequences. Consequently, with the goal changing, the old object file must be discarded. A new file must be opened, causing the extended looking time. However, the problem is that object files do not maintain any previous history of the objects (Treisman, 2006). They neither indicate anything about the identity of the object nor support object recognition (Carey & Xu, 2001).

Therefore, when only relying on the object file system, infants are not capable of distinguishing hand-reaching from rod-reaching, especially given that the two take the same path. This suggests that object file theory is incapable of explaining the extended looking time in studies that involved infants' understanding of the unobservable constructs.

Infant looking time paradigms are powerful behavioral paradigms to study infants' mental capacity. Object File Theory, for one, is a strong theoretical underpinning for the paradigms. But it is only so when the goal was to investigate infants' understanding of physical objects. This theory holds that the operations of object files, the basic processing unit of object-based visual attention, are the underlying causes for the differences in infants' looking time. This explanation only works within a rather limited scope, i.e., studies concerning infants' perception and representations of physical objects and their properties. When using the looking time paradigm to study infants' understanding of goal-directedness, intention or mental states, this theory becomes rather irrelevant. This is not to deny the possibility that infants can encode unperceivable features of the scenes. Infants may have a separate attentional mechanism that is dedicated to encoding aspects of visual scenes related to Theory of Mind, as Leslie (1994) proposed. But in the current context, our understanding of infants' visual attention cannot yet explain how a goal can be encoded in our representations of object files. If there does exist a system that encode unperceivable states like goal, then a question following is how the system coexists and interacts with the Object File system, which has a direct interface with perceptible features. In conclusion, for the Object File Theory to support the interpretations of the paradigms, it is crucial to delimit the explanatory scope from the beginning. Researchers interested in infants' understanding of the unperceivable properties should exercise extra caution in adopting the appropriate testing procedures and testing paradigms.

2.4 Dual Process Theory

Contemporary of Sokolov's Comparator Model, Dual Process Theory was first summarized and articulated by Groves and Thompson (1970). This theory originated from studies based on neural circuits and animal models (For example, in Groves and Thompson (1970) they used the acute spinal cat's hindlimb flexion reflex as an example). In the mid-1980s, this theory was incorporated into infant research. Researchers conducted a series of studies using the familiarization paradigm to test the predictions yielded by the Dual Process Theory (Bashinksy, Werner & Rudy, 1985; Kaplan, Scheuneman, Jenkins & Hilliard, 1988; Kaplan & Werner, 1986, 1987). At the time, Dual Process Theory was viewed as a rival theory to the Comparator Theory. They could yield conflicting predictions of infants' looking time pattern. Curiously, with the experimental evidence being consistent with the Dual Process Theory's prediction, this theory still remains rather isolated in the field and the conflict was left unaddressed by the proponents of the Comparator Theory. To be clear, Dual Process Theory is not forgotten. Review papers still acknowledge the contribution of this theory (Colombo & Mitchell, 2009; Kavšek, 2013; Oakes, 2010). Nevertheless, neither these reviewers nor the practitioners in the field addressed the conflicts between the two theories. It is the goal of this section to evaluate the Dual Process Theory in light of the original experiments conducted in the mid-1980s, as well as the theory's relations with the aforementioned theories.

2.4.1 Dual Process Theory: Habituation and Sensitization.

Dual Process Theory holds that there are two distinctive processes, sensitization and habituation, jointly determining the responses to the stimulus The sensitization process is an excitatory process that influences the organism's "state", which can be characterized by the level of arousal and activation; the habituation process is an inhibitory process that works specifically on the Stimulus-Response pathway in the organism. To put it differently, sensitization process influences some general level of activations in the organism or the systems of interests. In contrast, the habituation process works on the S-R pathway that is stimulus-specific, without imposing an impact on other systems or pathways. These two processes are species-general with neurophysiological underpinnings such as synaptic changes. Note that the characterizations of the State system and the S-R pathway are abstract. State system refers to "a construct of state, arousal, activation, excitability or responsiveness, etc" (Groves & Thompson, 1970). And the S-R pathway is any pathway that qualifies as "the most direct route through the central nervous system from stimulus to discrete motor response" (Groves & Thompson, 1970). It does not refer to any specific biological pathway implemented in any specific parts of the neural system.

The two processes differ in their trajectories. The sensitization process first leads to an increase in response, then followed by a decrease. In contrast, the habituation process results in a gradual decreasing response. This decrement develops exponentially and eventually reaches an asymptote. During an experiment, the responses observed by experimenter are the accumulations of the two processes's output. Depending on the properties of the stimulus, the two processes may give different outputs. For some stimuli, the increase driven by the sensitization process may dominate and cause the observed response to show an increase-first, decrease-second curve. On the contrary, if the habituation process dominates throughout, the observers would see a uniformly decreasing curve, though the rate of decrement will differ as to the sensitization process transit from increment to decrement..

The Dual Process Theory also views dishabituation differently. Early on, Thompson & Spencer (1966) observed that after showing decreased response to a repeated stimulus A, the organism would show an increased response to a novel stimulus B. This is normally known as dishabituation. But interestingly, after tested on B, when the organism is retested with the old stimulus A, the organism will also increase the response to the old stimulus A. It will "dishabituate" to the old stimulus. This phenomenon, the recovery of the response to the old stimulus, is named "Thompson-Spencer dishabituation". Dual-Process Theory regards both the dishabituation and the Thompson-Spencer dishabituation as the result of the sensitization process. Meanwhile, the habituation process is not disrupted. The theoretical characterizations of the increased response leads to predictions that are contradictory with the predictions yielded by the Comparator Theory. This contradiction will be further explored in the sections below.

2.4.2 Dual Process Theory in Infant Visual Behaviors

Infant researchers introduced The Dual Process Theory into the field via a series of studies using the Familiarization paradigm. At that time, Comparator Theory dominated the understanding of infants' visual behaviors (Cohen, 1972, 1973; Cohen, DeLoache, & Rissman, 1975). The consensus was that infants' decreased looking time toward a repeatedly shown stimulus was due to the decreasing discrepancy between the "mental representation" of the stimulus and the actual stimulus. However, researchers who were skeptical about the Comparator Theory proposed that Comparator Theory cannot explain a few anomalies from the looking time data: the nonmonotonicity of the decrease in looking time and the changes in Thompson-Spencer dishabituation. These anomalies, according to some researchers, were better explained by the Dual Process Theory.

2.4.1.1 Nonmonotonicity. The nonmonotonicity of the decrease in looking time refers to the phenomenon that infants' looking time during the pretest phase does not follow a decreasing linear trend. Instead, in the first couple of trials, they tend to look longer in each trial before their looking time starts to drop. Early on, research had shown that nonmonotonicity was found in the majority of the five-month-old and ten-month-old infants. Using the fixed-trial length procedure, McCall(1979) showed that five-month-olds looked the longest in the trials other than the first trial. This means that their looking time peaked in later trials in the pretest phase. At ten-months-old, more than 90% of the infants showed this pattern. Consistent with this finding, Bashinski and colleagues (1985) found this pattern is present in 4-month-olds. In this study, 4-month-olds infants saw a black and white checkerboard for eight trials, with each trial lasting 10 seconds. They found a nonmonotonic looking pattern. The fixation time from trial one and trial two increased, followed by a systematic decrement in the following trials. Comparator Theory would not predict this trend: if the longer looking time is driven by a larger discrepancy between the mental representation of the stimuli and the stimuli itself, then how could a continuously "representation-building" infant first have an increasing "discrepancy"? In contrast, the Dual-Process Theory explains this phenomenon well: the increased looking duration was due to the sensitization process.

The studies mentioned above were both run using the fixed trial length procedure, rather than the infant control procedure. These two procedures differ mainly in the determinations of the onset and offset of each trial and each phase. For the fixed trial length procedure, the experimenters predetermine the number of trials and the duration of each trial during the pretest phase. For the infant control procedure, each trial begins as the infant orients toward the screen and the trial ends when the infant looks away for more than a couple of consecutive seconds. The total number of trials in the pretest phase also depends on the infant. The experimenter will set a "habituation criterion" to determine when the pretest phase ends, and the test phase starts. For example, the pretest phase may end when the average of the infant's looking time in three consecutive trials falls below 50% of the first three consecutive trials. This infant control procedure is popular among practitioners of infant looking time paradigms. On the contrary, the fixed-trial length procedure was criticized as unable to accurately reflect infants' interests. For example, the infant might be still looking when the trial ended, or the infant might look away at the very beginning (Cohen, 1976; Bornstein, 1985).

However, this accusation is not entirely fair for the fixed trial length procedure. Infant control procedure is problematic too. First, the selection of the criterion for ending trials and pretest phase is rather arbitrary. The implicit assumption for selecting "maximal-look-away" time was that the infant has somehow "exhausted" her or his attention in this trial when she or he looks away for more than a predetermined duration of time. It still does not eliminate the possibility that the infant will return to the stimulus after 2 seconds. In the real world, the infant gets distracted and looks around all the time. If the goal for infant looking time paradigms is to probe the infants' learning process, then to artificially create a distraction-free looking time might endanger the ecological validity of the construct. The second problem with the Infant Control Procedure is that the procedure would inevitably mask some important aspects of the looking time, especially how the looking time changes from trial to trial naturally. The

Comparator Theory may hold that the trial-to-trial change is rather irrelevant: the infant will reach the "complete representation" as long as the infant meets the habituation criterion. Neither the total duration of looking nor the dynamics of the changes in looking were interesting for the Comparator Theory. Nevertheless, to know what really happens as the infant is repeatedly shown a stimulus, the fine-grained analysis of within-trial dynamics is indispensable. In this case, one cannot simply attribute the nonmonotonicity to an artifact created by the fixed trial length procedure. It risks overlooking an important feature of infant looking time dynamics.

2.4.1.2 Changes in Thompson-Spencer Dishabituation. Thompson-Spencer dishabituation refers to the recovering responses to an already-habituated stimulus (see figure 2d). A study illustrating Thompson-Spencer dishabituation often contains three phases. First, during the pretest phase, the infant would see a stimulus repeatedly. Then, a novel stimulus will appear. Following the novel stimulus, the old stimulus from before will reappear. According to the Dual Process Theory, the infant would look longer at the familiar stimulus during the third phase because of the sensitization process was activated by the novel stimulus. For this interesting phenomenon, unfortunately, the Comparator Theory never has any discussion. Maybe there is a reasonable explanation: once the infant has built the complete representation of the stimulus during the first phase, then any recovering of the response during the third phase is attributable to the decaying of the representations. This is nothing but a rather trivial observation. However, Dual Process Theory holds this recovery is theoretically relevant. Thompson-Spencer dishabituation provides an exciting opportunity to disentangle the two theories: how would the length of the second novel-stimulus phase influence the Thompson-Spencer dishabituation? For Comparator Theory, the long gap between the first phase and the third phase gives more time for the representation to decay. This will lead to a larger increased response, a larger Thompson-Spencer dishabituation. On the contrary, for Dual Process Theory, if such recovery is a product of the sensitization process, then the longer duration will also lead the response provided by sensitization process to decay. Therefore, the Thompson-Spencer dishabituation effect would be smaller.

Kaplan and Werner (1987) conducted an experiment to test these two rival theories. In this study, the 4-month-olds infants were assigned to three groups. The three groups differed in the numbers of novel stimulus presentations in the second phase. Infants in each group saw 6 trials, 4 trials, or 2 trials respectively. The rest of the testing procedure was identical across the three groups. They first saw eight trials of 4 x 4 black and white checkerboard. Then during the second phase, they were exposed to different number of trials of novel stimulus. In the third phase, they were tested with the stimulus they first saw during the first phase. What they found was consistent with the prediction by the Dual Process Theory. Infants who saw six trials in the second phase showed the least amount of response recovery in the third phase. This finding is impossible if there is only one single process going on when the infant is viewing the stimulus.

2.4.3. Relationship with other theories

Among the four theories discussed in this paper, Dual Process Theory is the one least known. Except for occasional mentions in review papers (e.g, Kavšek, 2013; Sirois & Mareschal, 2002; Colombo & Mitchell, 2009), the impact of this theory is minimal. A recent search in March, 2020 on Web of Science suggests that the paper published in 1987 by Kaplan and Werner titled “Sensitization and Dishabituation of Infant Visual Fixation” has only been cited 17 times. Historical coincidence aside, this statistic indicates that the Dual Process Theory remains relatively isolated from the mainstream in the field. Consequently, the advocates of Comparator Theory and Object File Theory have not entertained perspectives held by Dual Process Theorist. Neither do they address the conflicting interpretation offered by the Dual Process Theory. However, for theoretical clarity, it is still important to consider how these theories would relate to each other. This will be the goal of the current section.

One response to the nonmonotonicity is that the trend is not directly against the core of the Comparator theory. The trend is a feature of the behavioral responses. There are additional linking steps between the behavioral responses and the underlying mental representations. Other factors, such as the state-system proposed by the Dual Process Theory, or the operation of the Object File System, may both play a role in determining the ultimate behavioral output. But neither of these factors undermine the completion of the underlying mental representation. The building of mental representation can still follows a monotonic trend. Comparator Theorists and Object File Theorists can also point out that the Dual Process Theory operate on a different level, under Marr's framework. The Dual Process Theory is grounded at the implementational level, whereas the Comparator Theory is on the algorithmic level.

These potential responses are reasonable. They are to be congratulated for pointing out an important direction: if the inconsistency in interpretation reflects the theories operating on different levels, then the need to find the connections between the theories across levels is even stronger. Theory choice does not have to be mutually exclusive, one or the other. But one has to be clear about the explanatory scope of each theory, and the connections between different theories. How would the representations and the processes postulated by the Comparator theory and the Dual Process theory be reflected in the implementational level? To what extent can we extend the account in the implementational level to the algorithmic level? Is it possible, from the perspective of the implementational level account, to provide a better judgment on the explanatory scope of the Object File Theory and the Comparator Theory? These are future questions to be addressed by more empirical investigations.

Secondly, the Comparator Theorist and the Object File Theorist may also turn to the procedural differences. With Infant Control Procedure now dominating most of the infant looking time studies, the nonmonotonicity found in the Fixed Trial Length procedure should not be a concern at all for those who no longer used this procedure. Similarly, for the findings on the Thompson-Spencer habituation paradigm, the disconfirming evidence for the Comparator Theory perhaps also plays only a negligible role. After all, the three widely used infant looking time paradigms typically only involve two main phases, a pretest phase and a test phase. Infants are rarely, as the infants in Kaplan and Werner's study, exposed to a third "Thompson-Spencer dishabituation" phase following the test phase. So how would the discrepancy in the predictions and findings influence the conception of looking time paradigms at all? As long as the Comparator Theory can offer a reasonable account for what is going on in the pretest phase and test phase, what else should we expect from the theory?

However, this attitude toward the Dual Process Theory is misleading. It raises two more challenges. First, it is still unclear what is causing the different patterns across the two testing procedures. The end goal for the infant looking time paradigms is to use looking patterns to infer some properties of the underlying psychological processes. If we find systematic differences in looking patterns when different procedures are being used, then what are the factors causing the difference? Are the usage of attention-getter and the trial-ending criteria influencing the overall information processing performance of the infants? By systematically manipulating the length of each fixed-length trial, is it possible to change the trend of the nonmonotonicity? As for the Thompson-Spencer dishabituation effect, it should not be considered as something limited to the usage of a three-phase study. According to Colombo and Mitchell (2009), the order in which the test stimuli are presented may lead to the “contamination” of Thompson-Spencer dishabituation. For example, assuming that during the pretest phase, infant A was repeatedly shown the stimulus X. During the test phase, two types of stimuli were presented: X’ and Y. X’ was designed to be the “familiar” stimulus, more similar to the original familiarized stimulus X than the stimulus Y. Assuming that in the test phase the infants were always shown the stimulus Y before stimulus X’. Now, if the infant looks longer at the stimulus X’, we can no longer claim that the longer looking time is due to a “familiarity preference” or, be certain that the “novelty preference” is absent. The precedence of novel stimulus Y had already sensitized the infant before the familiar stimulus X being presented. A standard practice in the field is to randomize or semi-randomize the order of the stimulus. Consequently, this leads infants to be subject to different levels of Thompson-Spencer dishabituation. An infant who sees three trials of “novel stimulus” before one trial of “familiar stimulus” must-have undergone a different degree of sensitizations than an infant who sees only one trial of “novel stimulus”. However, most of the researchers average across the looking time in “novel trials” and “familiar trials”. It is difficult to assess how the sensitization process may cause the specific effect to be overestimated or underestimated. Consequently, the presence of Thompson-Spencer dishabituation introduces more noise into the ambiguous looking time data.

One cannot dodge the influence of the Thompson-Spencer effect by always presenting the familiarity stimulus before the novelty stimulus during the test phase. Doing this will introduce the Order Effects instead. Thus, if the Thompson-Spencer dishabituation is not avoidable through experimental design, then a priority for researchers is to sketch out a more precise framework to analyze the sensitization effect of the stimuli based on perceptual features. Some have investigated how the specific perceptual features of the stimulus can lead to different levels of arousal in infants. Notably, Kaplan and colleagues (1986) investigated how the contrast of the visual patterns might influence the sensitization process in infants. They took a psychophysical approach to show how spatial frequency analysis is relevant to the dynamics of infants' looking behaviors. For the field moving forward, if the goal is to have a well-grounded understanding of infants' visual behavior, then more works like this are needed to supplement further details on how different perceptual features can influence the looking behaviors.

Finally, the connection between Multifactor Model and Dual Process Theory needs more exploration. The three factors, age, task complexity and the familiarization time all may influence the infant's behaviors under the framework of Dual Process Theory. Does the process of sensitization and habituation have different rates of maturation during infancy? How would age and familiarization time alter the trajectory of the two processes' influence on infants looking behavior? Is the same level of complexity being weighed differently by the two processes? Is it possible that a given level of complexity may influence the sensitization process more than the habituation process, or the other way around? All of these questions could potentially provide important insights for researchers treating looking time data as an important way to probe into cognitive development.

Dual Process Theory deserves more attention. By postulating two independent and simultaneous processes, sensitization and habituation, the theory reveals some overlooked details of infants' looking behaviors. One is a non-monotonic decrease in looking time, which is shown when using the Fixed Trial Length Procedure. The other is the Thompson-Spencer dishabituation, which is shown in an additional phase after the infants have "dishabituated" to the novel stimulus. Although the testing procedure giving rises to these findings deviates from the mainstream infant research nowadays, it still teaches a valuable lesson to us on the important influences from details in experimental practices.

3. Concluding Remarks

Looking time paradigms are central to developmental psychology. In the first two years of their life, infants undergo rapid changes in their cognitive capacity. These babies carry important clues to the nature of human mind, yet their behavioral repertoire is so limited that it imposes great challenges to the researcher. Looking time, as one of a few measurements available to collect from infants at birth, has provided important insights on the underlying mechanism for development. On the other hand, however, looking time data is often challenging to interpret. There is a great deal of underdetermination between the differences of the looking duration and the factors driving the looking time. The goal of this paper was to provide an overview of different theoretical accounts proposed in the field, and to see if the theories are compatible with each other.

In this paper, four theories have been surveyed: the Comparator Theory, the Multifactor Model, the Object File Theory, and the Dual Process Theory. These four theories were selected due to their prevalence in the field, suggesting that they are all relatively well-received theories convincing many. As stated in the beginning, each theory should be able to account for the looking time pattern in the pretest phase and the test phase. During the pretest phase, the theories should explain the decreased looking time across the trials; During the test phase, the theories should be able to predict and account for three possible outcome: novelty preference, familiarity preference and null preference. While all but Multifactor Model provide a possible mechanistic explanation for the looking time curve, each theory offers a slightly different explanation for the looking time pattern. The Comparator theory holds that it is caused by the discrepancy between built mental representation and the physical stimuli; Object File Theory focuses solely on the performance in the test phase, interpreting the differences in looking time as due to the mismatch in the number of object files in operation; Dual Process Theory attributes the rise and fall in the looking time to two separate processes, sensitization and habituation, working simultaneously. In contrast with these three theories, Multifactor model, on the other hand, is postulated on an interpretation level. Three factors work collectively to determine the nature of the prolonged looking time: age, familiarization time, and task complexity. Although this model does not hint at the causal connections between these three factors and the looking behaviors, it is still valuable in systematizing the possible inferences made based on the looking time patterns. Moreover, the Multifactor Model also directly hints at the influence of test design on the effect of looking duration.

This paper is set to answer whether these theories are compatible with each other. If compatible, then the unified account can provide clearer guidelines for interpreting the looking time data. If incompatible, then pinpointing the conflicts between theories will be helpful for generating more further research insights.

How do these theories relate to each other? The answers are mixed. Among all the theories, Comparator Theory is the one with widest acceptance. Overall, Comparator Theory is compatible with the Object File Theory. Both Comparator Theory and the Object File Theory emphasize on the role of comparison in learning. For Comparator Theory, the comparison is between the built mental representation and the physical stimuli. For Object File Theory, the comparison is between different numbers of operating object files. At the current stage of the two theories, the actual mechanism for comparison and the underlying processes for comparison to happen remain unexplained. It is worth noting that there is a separate line of research focusing on the important role of comparison in learning, starting from early infancy throughout childhood and continue to be important during adulthood. Research has shown evidence that visual comparison can facilitate analogical abstraction in 3-month-olds (Anderson, Chang, Hespos, & Gentner, 2018). It also aids categorizing ability and promoting false-belief reasoning in young children (Kotovsky & Gentner, 1996; Hoyos, Horton, & Gentner, 2015). For adults, comparison is also shown to facilitate analogical transfer when combined with explanation (Edwards, Williams, Gentner & Lombrozo, 2019). However, the usage of "comparison" in these studies often refers to its operationalized definition. The tasks are designed to make the participants compare, either explicitly as in the case of adult study, or implicitly as in the case of infant studies on visual comparison. Therefore, it would be an interesting direction to explore the connections between the two ways of conceptualizing comparison. To what extent is the "comparison" in the Comparator Theory and Object File Theory be held accountable for the facilitation effect on learning across age groups? What is the scope and limitation of the learning by comparison? In what cases would comparison be detrimental to learning? How does the facilitation provided by comparison connect to other cognitive factors such as executive function? These are all questions left unexplored. But they are crucial for us to understand the theoretical underpinning of the infant looking time paradigm.

In contrast, the relationship between Comparator Theory and Dual Process Theory is far more incompatible. Their assumptions about the underlying mechanism in the changes in looking time differ. While Comparator Theory holds that there is only one process driving the looking time fluctuation, the Dual Process Theory assumes that there are two distinct processes, the habituation process and the sensitization process, working simultaneously. As discussed early on, different assumptions lead to different predictions. Comparator Theory would not have predicted the nonmonotonicity of looking time pattern, nor would it have predicted the increase in Thompson-Spencer dishabituation when the interval between test and retest is extended. Unfortunately, given the scarcity of literature on the Dual Process Theory in the context of infant looking time paradigms, this contradiction is difficult to tease apart without more empirical data and more refined measurement. It might be the case that these theories are not completely incompatible — two processes work concurrently, with the Comparator Theory capturing the habituation process as postulated by the Dual Process Theory. It might be very difficult to have an empirical grasp on the two competing theories with the blunt instrument, looking time. Therefore, to further explicate the relationships between the two theories, one should consider consulting other measurements, such as other behavioral measurement or neurophysiological measurement, that can help sharpen predictions.

The relationship between Dual Process Theory and the Object File Theory suffers from similar uncertainty. From the perspective of Marr's framework, these two theories operate on different levels of analysis: Dual Process Theory is closer to the implementational level, whereas Object File Theory is closer to the algorithmic level. The current available empirical data is yet to generate a sufficiently sound account to connect the theory from the two ends. Thus, future work is needed to explicate the connection between the two theories.

These summaries might be disappointing: after more than half a century of using looking time as measurement in infant research, we are still in need of a unified and convincing theory to justify the measurement and the paradigms using this measurement. Maybe it is more than a mere coincidence that all three mechanistic theories reviewed here came from a different subfield — the Comparator Theory and Dual Process Theory originated from neurophysiological studies investigating learning theory, and the Object File Theory was inspired by the progress made in adult object-based attention. Borrowing theories from other domains can be beneficial at the initial stage of the field. But eventually, the field must be able to synthesize and generate new theories adapted to the specific cohort of problems faced by the field. Looking time as a measurement, as we have seen, is not informative enough to enable theory-building researchers to have a full grasp of its underlying mechanisms. Alternatively, a growing number of researchers have turned to different measurements in search of converging evidence for the theory advocated: reaching behavior was used to support findings in Violation of Expectation studies (Hespos & Bailargeon, 2008); the analogy between adults' decreasing neural response toward repeated stimuli in fMRI and infants' decreasing looking time has been explored (Turk-Browne, Scholl & Chun, 2008); other more refined visual measurement, such as predictive looking, also drew more and more attention in recent years (Daum, Attig, Gunawan, Prinz & Gredebäck, 2012).

Our understanding of looking time can be aided by coupling it with other measurements. But this does not mean that we should abandon the looking time, or simply put it into the attic. Although looking time is a blunt instrument, there is still an available starting point for more refined theory building: the Multifactor Model Framework. In contrast with the other three theories, Multifactor Model does not directly explain the underlying mechanisms for the looking time patterns. Yet it has provided abundant guidance in interpreting the looking time data. From a practical perspective, this model is the most valuable one in terms of guiding research design and interpretation. Nevertheless, the current three factors articulated by this theory remain rather vague: age, familiarization time and complexity. They all lack a sufficiently precise definition. What exactly is the developmental trajectory behind the age factor? How would different procedures, i.e., infant control procedure versus fixed trial length procedure, influence the familiarization time factor? And finally, how do we measure complexity in infancy? Early work simply used numbers of toys in an array as a proxy for complexity (Hunter, Ames,& Koopman, 1983). Recent work has used predictability as a proxy for complexity, showing that infants prefer stimuli that are neither too simple nor too complex (Kidd, Piantadosi, & Aslin, 2012). These are very different ways to operationalize complexity. How are they connected to each other? Is it possible to find a unified way to capture this construct? There are many questions can to be asked on each of the factor and the interaction between each factor. It is not difficult to imagine that a deeper understanding of these factors would help us make inferences about the underlying mechanism in turn. Unlike the yet to be done works awaiting the other theories, Multifactor Model is the one most promising to yield precise quantitative predictions.

While it is almost guaranteed that we will be able to obtain more refined behavioral and neural measurement in the future, at no point should the technological progress excuse us from developing the theoretical foundations of looking time paradigms. Science is a cumulative endeavor. Much of what we presently know about infant perception and cognition stemmed from these three infant looking time paradigms: familiarization, habituation and Violation of Expectation. It is never too late to know better about how we get to know what we know.

References

Anderson, E. M., Chang, Y. J., Hespos, S., & Gentner, D. (2018). Comparison within pairs promotes analogical abstraction in three-month-olds. Cognition, 176, 74-86.Aslin, R. N. (2007). What's in a look?. Developmental science, 10(1), 48-53.

Bashinski, H. S., Werner, J. S., & Rudy, J. W. (1985). Determinants of infant visual fixation: Evidence for a two-process theory. Journal of Experimental Child Psychology, 39(3), 580-598.

Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. Developmental psychology, 23(5), 655.

Baillargeon, R. (2004). Infants' physical world. Current directions in psychological science, 13(3), 89-94.

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. Child development, 62(6), 1227-1246.

Baillargeon, R., & Graber, M. (1987). Where's the rabbit? 5.5-month-old infants' representation of the height of a hidden object. Cognitive Development, 2(4), 375-392.

Baillargeon, R., Li, J., Luo, Y., & Wang, S. H. (2006). Under what conditions do infants detect continuity violations. *Processes of change in brain and cognitive development: Attention and performance XXI*, 163-188.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. Cognition, 20(3), 191-208.

Biro, S., & Leslie, A. M. (2007). Infants’ perception of goal-directed actions: development through cue-based bootstrapping. Developmental science, 10(3), 379-398.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set× event set design. Developmental psychology, 33(3), 408.

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis.

Burke, L. (1952). On the tunnel effect. Quarterly Journal of Experimental Psychology, 4(3), 121-138.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: Beyond object files and object tracking. Cognition, 80(1-2), 179-213.

Carey, S. (2002). Evidence for numerical abilities in young infants: a fatal flaw?. Developmental Science, 5(2), 202-205.

Christodoulou, J., Lac, A., & Moore, D. S. (2017). Babies and math: A meta-analysis of infants' simple arithmetic competence. Developmental psychology, 53(8), 1405.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. Annual review of psychology, 62, 73-101.

Clearfield, M. W., & Mix, K. S. (1999). Number Versus Contour Length in Infants' Discrimination of Small Visual Sets. Psychological Science, 10(5), 408–411.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, 869-879.

Cohen, L. B. (1973). A two process model of infant visual attention. Merrill-Palmer Quarterly of Behavior and Development, 19(3), 157-180.

Cohen, L. B. (1976). Habituation of infant visual attention. Habituation: Perspectives from child development, animal behavior, and neurophysiology, 207-238.

Cohen, L. B., DeLoache, J. S., & Rissman, M. W. (1975). The effect of stimulus complexity on infant visual attention and habituation. Child Development, 611-617.

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. Developmental science, 5(2), 186-201.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. Neurobiology of learning and memory, 92(2), 225-234.

Colombo, J., Richman, W. A., Shaddy, D. J., Follmer Greenhoot, A., & Maikranz, J. M. (2001). Heart rate-defined phases of attention, look duration, and infant performance in the paired-comparison paradigm. *Child Development*, *72*(6), 1605-1616.

Daum, M., Attig, M., Gunawan, R., Prinz, W., & Gredebäck, G. (2012). Actions seen through babies' eyes: A dissociation between looking time and predictive gaze. Frontiers in Psychology, 3, 370.

Edwards, B. J., Williams, J. J., Gentner, D., & Lombrozo, T. (2019). Explanation recruits comparison in a category-learning task. Cognition, 185, 21-38.

Feldman, J. Formation of visual "objects" in the early computation of spatial relations. Perception & Psychophysics 69, 816–827 (2007).

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. Cognitive psychology, 59(2), 154-179.

Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological bulletin*, *65*(5), 305.

Green, E. J., & Quilty-Dunn, J. (2017). What is an object file?. The British Journal for the Philosophy of Science.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review*, *77*(5), 419.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly?. Infant behavior and development, 21(2), 167-179.

Hespos, S. J., & Baillargeon, R. (2001). Infants' knowledge about occlusion and containment events: A surprising discrepancy. Psychological Science, 12(2), 141-147.

Hespos, S. J., & Baillargeon, R. (2008). Young infants' actions reveal their developing knowledge of support variables: Converging evidence for violation-of-expectation findings. Cognition, 107(1), 304-316.

Hofstadter, M., & Reznick, J. (1996). Response Modality Affects Human Infant Delayed-Response Performance. *Child Development, 67*(2), 646-658. doi:10.2307/1131838

Hoyos, C., Horton, W., & Gentner, D. (2015). Analogical comparison aids false belief understanding in preschoolers. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), Proce

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. Developmental Psychology, 19(3), 338.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. Advances in infancy research.

Julesz, B. (1990). Early vision is bottom-up, except for focal attention. In Cold Spring Harbor symposia on quantitative biology (Vol. 55, pp. 973-978). Cold Spring Harbor Laboratory Press.

Kagan, J., & Lewis, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly of Behavior and Development*, *11*(2), 95-127.

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

Káldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what'and 'where'information. Developmental Science, 6(3), 360-373.

Kaplan, P., Scheuneman, D., Jenkins, L., & Hilliard, S. (1988). Sensitization of infant visual attention: Role of pattern contrast. Infant Behavior and Development, 11(3), 265-276.

Kaplan, P. S., & Werner, J. S. (1986). Habituation, response to novelty, and dishabituation in human infants: Tests of a dual-process theory of visual attention. Journal of experimental child psychology, 42(2), 199-217.

Kaplan, P. S., & Werner, J. S. (1987). Sensitization and dishabituation of infant visual fixation. Infant Behavior and Development, 10(2), 183-197.

Kavšek, M. (2013). The comparator model of infant visual habituation and dishabituation: Recent insights. *Developmental Psychobiology*, *55*(8), 793-808.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. PloS one, 7(5).

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. Child Development, 67(6), 2797-2822.

Leslie, A. M. (1995). Pretending and believing: Issues in the theory of ToMM. Cognition on cognition, 193-220.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'.

Trends in cognitive sciences, 8(12), 528-533.

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality?. Cognition, 25(3), 265-288.

Marr, D. (1982), Vision: A Computational Approach, San Francisco, Freeman & Co.

Newman, G. E., Choi, H., Wynn, K., & Scholl, B. J. (2008). The origins of causal perception: Evidence from postdictive processing in infancy. Cognitive psychology, 57(3), 262-291.

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. Journal of Cognition and Development, 11(3), 255-268.

Oakes, L. M., & Madole, K. L. (2000). The future of infant categorization research: A process-oriented approach. Child Development, 71(1), 119-126.

Piaget, J. (1954). The construction of reality in the child (M. Cook, Trans.). New York, NY, US.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition, 32(1), 65-97.

Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. Cognition, 50(1-3), 363-384.

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. Behavioral and brain sciences, 22(3), 341-365.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial vision, 3(3), 179-197.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. Perception, 22(4), 463-475.

Richards, J. E., & Casey, B. J. (1991). Heart rate variability during attention phases in young infants. *Psychophysiology*, *28*(1), 43-53.

Rivera, S. M., Wakeley, A., & Langer, J. (1999). The drawbridge phenomenon: Representational reasoning or perceptual preference?. Developmental psychology, 35(2), 427.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review*, *24*(1), 74-100.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*(5), 704.

Rubio-Fernández, P. (2019). Publication standards in infancy research: Three ways to make Violation-of-Expectation studies more reliable. *Infant behavior and development*, *54*, 177-188.

Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. What is cognitive science, 26-73.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. Cognitive psychology, 38(2), 259-290.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. Trends in cognitive sciences, 4(8), 299-309.

Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). Cognitive development, 10(2), 253-269.

Sirois, S., & Mareschal, D. (2002). Models of habituation in infancy. Trends in Cognitive Sciences, 6(7), 293-298.

Sirois, S., & Mareschal, D. (2004). An interacting systems model of infant habituation. Journal of cognitive neuroscience, 16(8), 1352-1362.

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. *Psychological review*, *106*(2), 235.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual review of physiology*, *25*(1), 545-580.

Spelke, E. S. (1988). Where perceiving ends and thinking begins: The apprehension of objects in infancy. In A. Yonas (Ed.), Perceptual development in infancy. Minnesota Symposium on Child Psychology (Vol. 20). Hillsdale, NJ: Lawrence Erlbaum, Assoc.

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. British Journal of Developmental Psychology, 13(2), 113-142.

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. Science, 210(4473), 1033-1035.

Thompson, R. F. (2009). Habituation: a history. *Neurobiology of learning and memory*, *92*(2), 127.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. Psychological review, 73(1), 16.

Treisman, A. (2006). How the deployment of attention determines what we see. Visual cognition, 14(4-8), 411-443.

Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. Frontiers in human neuroscience, 2, 16.

Wang, S. H., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. Cognition, 93(3), 167-198.

Wetherford, M. J., & Cohen, L. B. (1973). Developmental changes in infant visual preferences for novelty and familiarity. *Child Development*, 416-424.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. Cognition, 69(1), 1-34.

Woodward, A. L., Sommerville, J. A., Gerson, S., Henderson, A. M., & Buresh, J. (2009). The emergence of intention attribution in infancy. Psychology of learning and motivation, 51, 187-222.

Wynn, K. (1992). Addition and subtraction by human infants. Nature, 358(6389), 749-750.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. Cognitive psychology, 30(2), 111-153.

Appendix

Table 1

Summary of the Testing Procedures in Infant Looking Time Paradigms

	Phases	Procedure	Presentation	Stimuli*
Familiarization	Pretest - test	Fixed Trial Length	Pretest: both serial and simultaneous presentation Test: both serial and simultaneous presentation	Simple, Static
Habituation	Pretest - test	Infant Control Procedure	Pretest: both serial and simultaneous presentation Test: both serial and simultaneous presentation	Can be either simple and static or complex and dynamic
Violation of Expectation	Pretest - test	Infant Control Procedure	Pretest: Serial presentation only Test: both serial and simultaneous	Complex, dynamic

*Note that in the last column the nature of the stimuli is not constrained by theoretical necessity. There is no theoretical constraints on why some types of stimuli can only be used in one paradigm but not the other. Rather, it is primarily due to conventions in experimental practices.

Figure 1

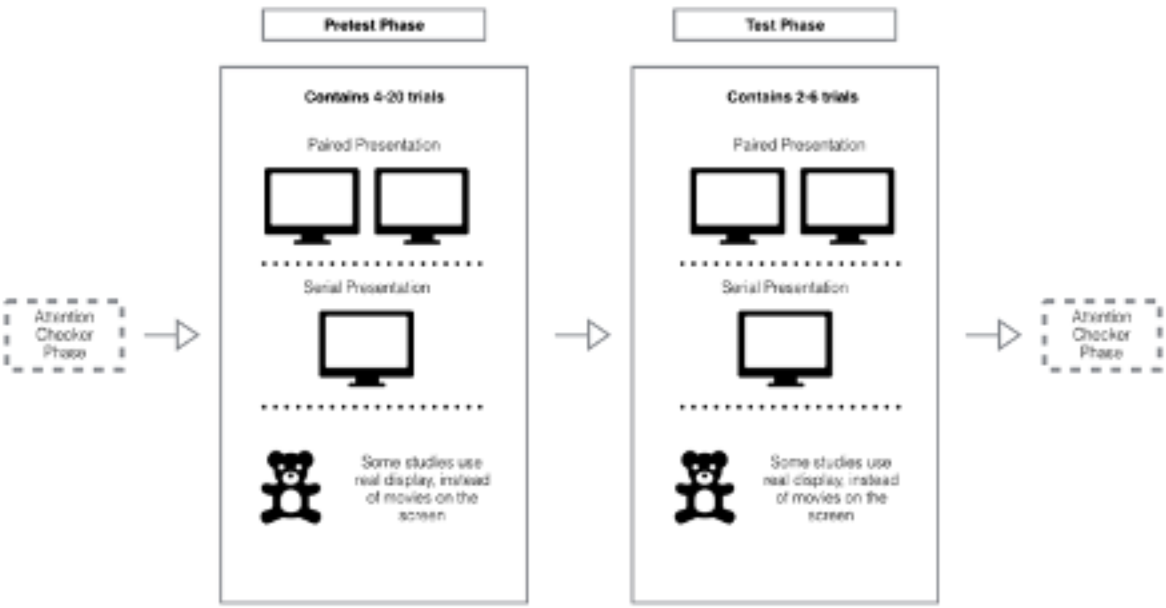
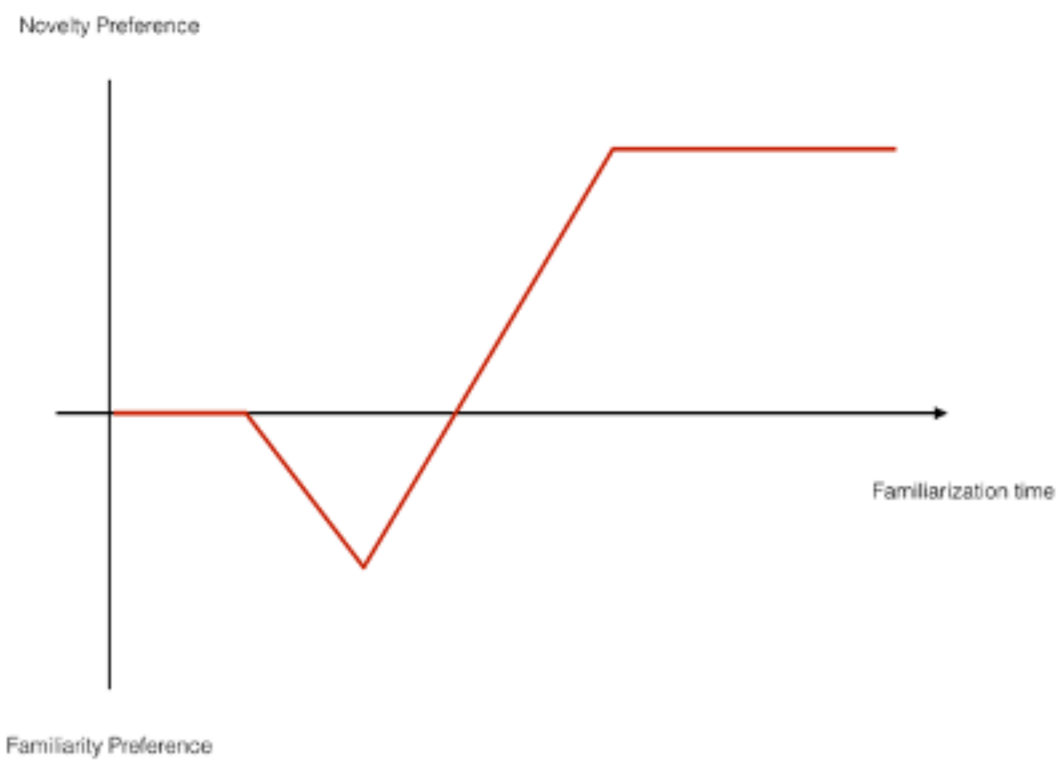


Figure 1. Each paradigm involves at maximum four different phases, but each phase can be implemented differently. The first and the last attention checker phases are designed to measure the baseline attention level. The content of the two attention-checker phase should be identical with each other but irrelevant to the pretesting and testing materials. During the pretest and the testing phase, studies diverge in the implementation and procedures, which lead to many debates about the interpretations.

Figure 2a

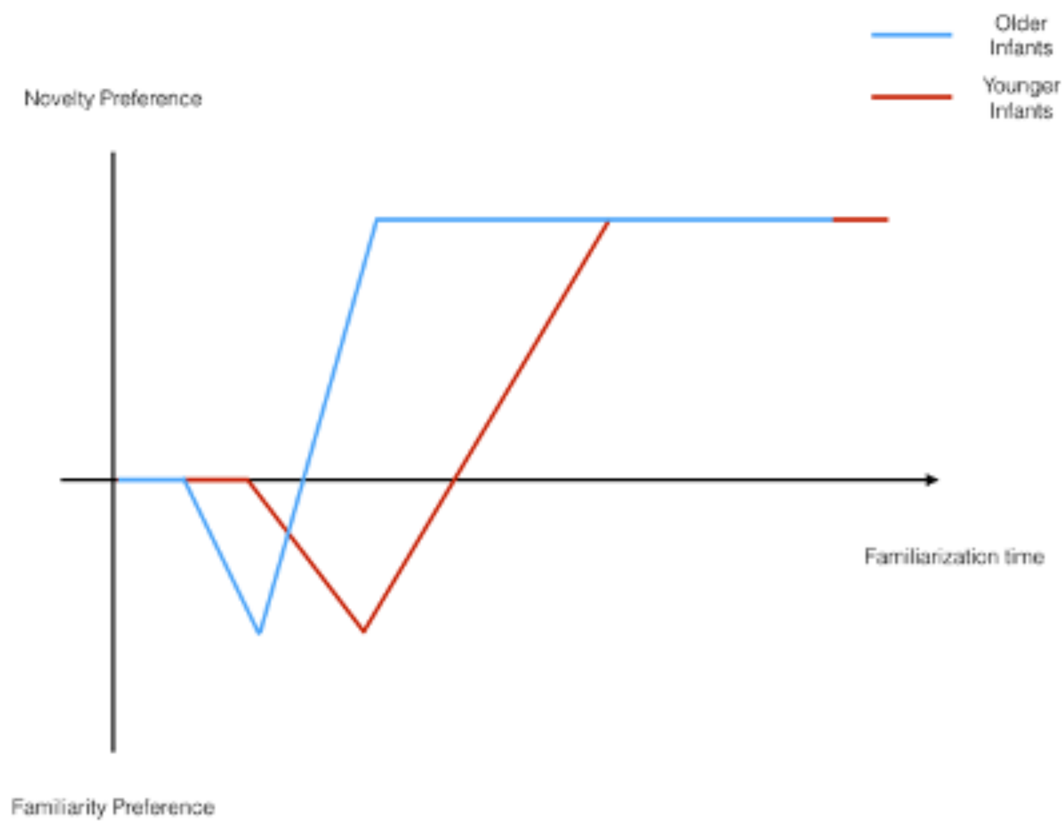
Preference shift in Infant Looking Time Paradigms



Adapted from Hunter & Ames (1988).

Figure 2b

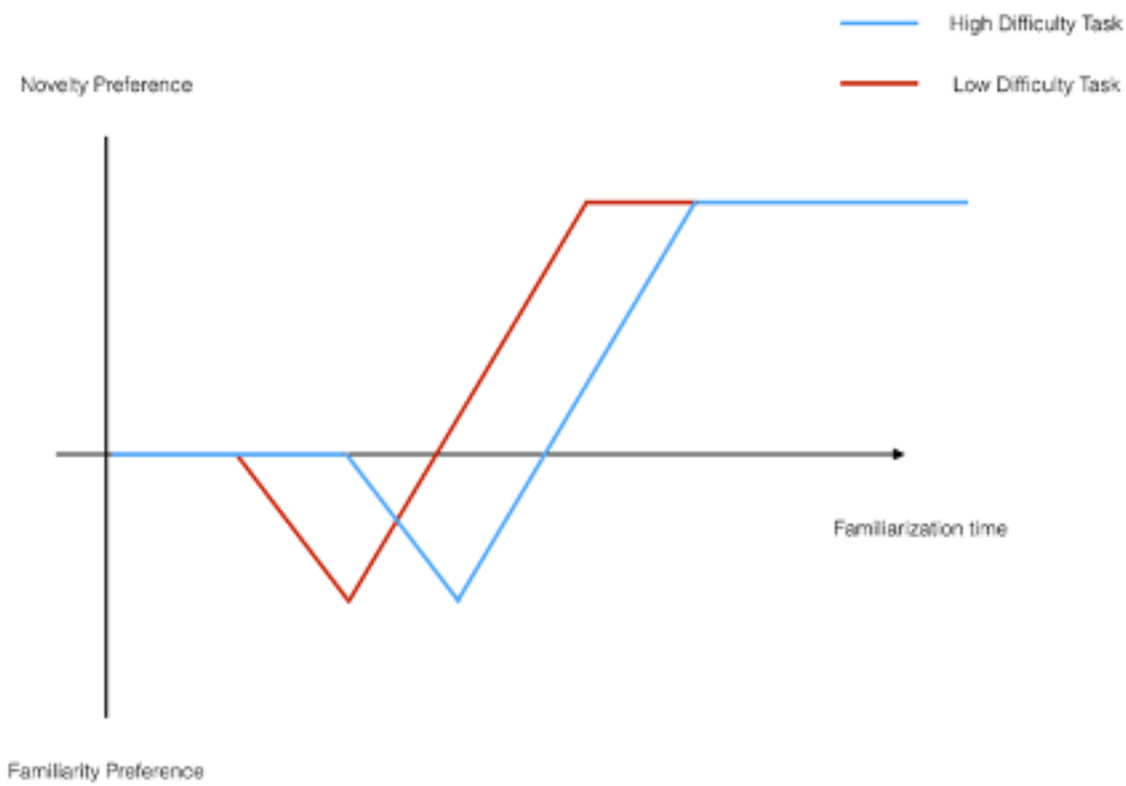
Preference shift in Infant Looking Time Paradigms, with the influence of age



Adapted from Hunter & Ames (1988)

Figure 2c

Preference shift in Infant Looking Time Paradigms, with the influence of task difficulty



Adapted from Hunter & Ames (1988)

Figure 2d

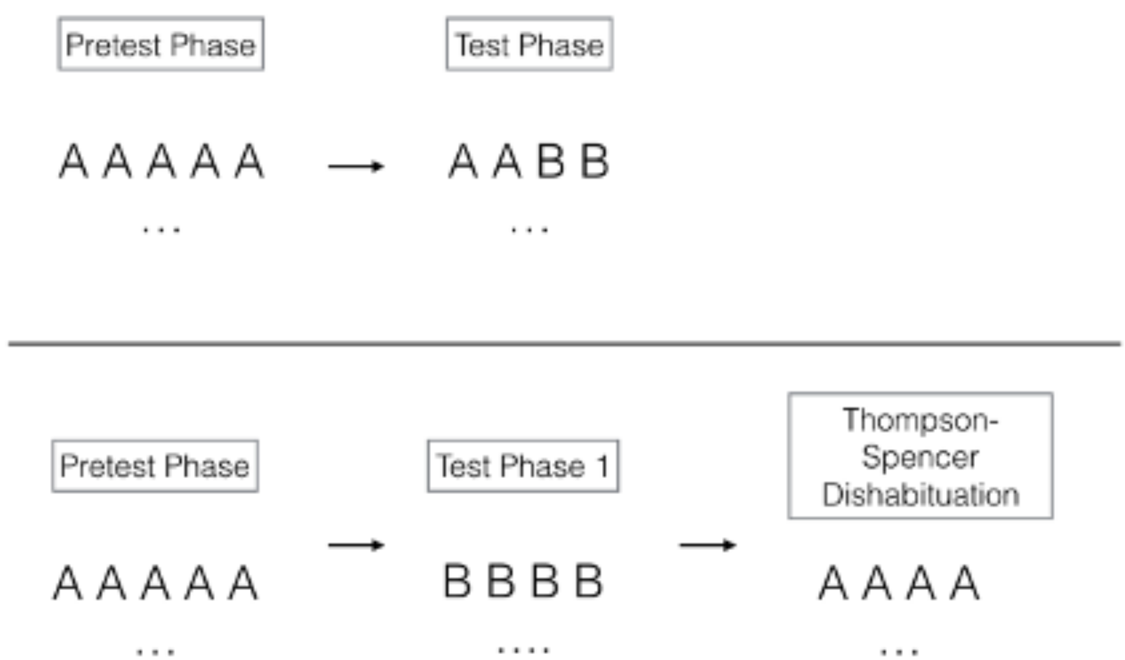


Figure 2d. *Thompson-Spencer dishabituation phase occurred after the test phase, in which the stimuli during Pretest Phase were presented again. The infants would show a recovery in looking time toward the already habituated stimulus during Thompson-Spencer Dishabituation phase.*