

SETTING BOUNDARIES: CHILDREN'S NEURAL AND
BEHAVIORAL EVENT COGNITION IS
ROBUST BUT STILL DEVELOPING
IN EARLY CHILDHOOD

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Susan L. Benear
August 2022

Examining Committee Members:

Nora S. Newcombe, Advisory Co-Chair, Psychology Department
Ingrid R. Olson, Advisory Co-Chair, Psychology Department
Vishnu P. Murty, Psychology Department
Jason M. Chein, Psychology Department
Susan Perlman, External Member, Washington University in St. Louis
Youssef Ezzyat, External Member, Wesleyan University

ABSTRACT

Segmenting our ongoing experience into events is a fundamental aspect of human cognition and memory. Prior work in adults has shown that we naturally and spontaneously parse our experience along event boundaries, both behaviorally and neurally, and that this is reflected in event memory as well. With this project, we aimed to examine whether children ages 4-7 also spontaneously track events while encoding naturalistic stimuli, whether they can behaviorally demarcate event boundaries, and how this influences their memory for events. Our results indicate that children can segment naturalistic stimuli into events like adults, but they do so with more variability, and their boundaries differ from adults' boundaries in terms of both location and consistency. Children's behaviorally-delineated boundaries were also reflected in a separate group of children's neural data, when examined both from a hypothesis driven and a data driven approach, indicating that young children's brains track events during perception. Last, we found that children's event segmentation grows more adult-like across early childhood, and that children who segment events more like adults may have better memory for those events. Overall, this study suggests that children's event cognition and memory is robust even at very young ages, but that it is still developing across early childhood and becomes more adult-like as children age.

I would like to dedicate this dissertation to my mother.

She has helped steer me throughout my life
with unwavering support through both trials and triumphs.

She is intelligent, introspective, and interesting,

and has always made me believe I could do

anything I wanted to...

even earn a doctorate.

ACKNOWLEDGMENTS

I would first like to acknowledge my mentors, Dr. Ingrid Olson and Dr. Nora Newcombe. They guided and encouraged me during my entire graduate experience, and I am proud to have worked with two such brilliant women in science—I will strive to emulate them throughout my career. I would also like to recognize Dr. Deepu Murty for his guidance on imaging analysis and his persistent positivity, and Dr. Susan Perlman for providing me with excellent neuroimaging data when I could not collect any myself. Next, I want to thank my lab mates Haroon Popal, Busra Tanriverdi, Katie Jobson, and Dr. Samantha Cohen, as well as Sean Zheng, who provided me with extensive coding and analysis aid, using their precious free time to assist me countless times. Thanks also to Drs. Chris Baldassano, Jeff Zacks, Janice Chen, David Smith, and Tom Olino, all of whom provided me with expertise regarding optimal design and analysis of this project. Additionally, I could not have recruited and tested all my participants without the help of my dedicated undergraduate research assistants, Sonya Afshan-Paroya, Jude Baffoe-Bonnie, Meghan Laties, and Obinnaya Onwukanjo, who all have bright futures ahead of them. I am exceedingly grateful to all of these people for the time they put in to helping this project come to fruition.

Last, I would be remiss not to acknowledge my loving and endlessly optimistic partner, Ross Holden, my mom and biggest supporter, Dr. Jennifer Myers, the rest of my family—especially my dad, Jay, my step mom, Laura, my sister, Amelie, and my brother, Cole—as well as my close friends, who have all shown me love, curiosity, and encouragement through my lengthy pursuit of a PhD. I truly could not have done this without you all.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
DEDICATION.....	iii
ACKNOWLEDGMENTS	iv
LIST OF FIGURES	vii
 CHAPTER	
1. INTRODUCTION	1
Background.....	1
Event Cognition	1
Event Memory	2
Development of Event Cognition and Memory	4
Motivation for the Current Study.....	6
2. METHODS	9
Participants.....	9
Stimuli.....	10
Event Segmentation Task	10
Episodic Memory Tasks	11
Working Memory and Verbal Intelligence Control Measures	11
Behavioral Procedure.....	12
Encoding and Control Tasks.....	12
Mnemonic Tasks.....	13

Event Segmentation Task	15
Imaging Data Acquisition and Preprocessing	18
Data Analysis	19
Behavioral Data	19
Imaging Data.....	22
3. RESULTS	27
Event Segmentation Agreement	27
Qualitative Data Exploration	27
Agreement Scores	29
Bootstrapping Analysis	30
Full Behavioral Models.....	32
Temporal Order.....	32
Recognition	34
Segmentation Agreement.....	37
Verbal IQ	38
Representational Similarity Analysis.....	39
Adult-Generated Boundary Results	39
Child-Generated Boundary Results	42
Hidden Markov Model.....	43
4. DISCUSSION	48
REFERENCES CITED.....	56

LIST OF FIGURES

Figure	Page
1. Across-boundary vs. within-event pairs.....	14
2. Plots of event segmentation across the entire television show.....	28
3. Event segmentation agreement.....	30
4. Temporal order memory.....	33
5. Recognition memory.....	36
6. Across-boundary and within-event comparisons used in the RSA.....	40
7. Pattern dissimilarity in EVC for adult-delineated boundaries.....	41
8. Pattern dissimilarity in PCC, LHC, LAG, and RAG for child-delineated boundaries	43
9. The 10 boundaries generated by the HMM for each ROI.....	45
10. The optimal number of boundaries for each ROI.....	46

CHAPTER 1

INTRODUCTION

Background

Event Cognition

As we take in perceptual information from our environment, we continuously process the input in myriad ways. Our sensory cortices make sense of light and sound, and higher-level regions integrate this information to build what we interpret as our experience. One way that we construct meaning from our experience is by breaking it into events. Humans do this naturally, even when not explicitly asked to do so (reviewed in Shin & DuBrow, 2021 and Zacks, 2020), and many authors theorize that event segmentation is evolutionarily adaptive. For one, demarcating an event boundary offloads processing of currently occurring episodes from working memory into long-term memory and updates the model for the next event, potentially with new predictions (reviewed in Zacks, 2020). Additionally, segmenting one's experience into events creates a narrative structure for episodic recall, which benefits mnemonic performance (Zacks, 2020).

In order to consider event segmentation, one must have a working definition of what an event is. Events can be “broadly defined as units of activity with an identifiable beginning and end,” (Shin & DuBrow, 2021). An event is something that “unfolds within a spatiotemporal framework and includes entities such as people and objects and the relations among them,” (Zacks, 2020). To summarize these definitions, events are bound by location and time, involve interactions and relationships between various aspects of our environment, and have an identifiable beginning and end. This could be applied to

something with a long timescale, such as a day at school, as well as those with a shorter timescale, like completing a worksheet in math class. Both of these occurrences would satisfy the requirements of being an event, yet one is on a “coarse” scale and one on a “fine” scale. Participants in the lab can easily shift between segmenting experience into more coarse- or fine-grained events simply by being instructed to do so (Magliano & Zacks, 2011), and there is evidence from neuroimaging studies that events are organized hierarchically in the brain as well (Baldassano et al., 2017; Geerligs, Van Gerven, Campbell, & Güçlü, 2021; Lee, Aly, & Baldassano, 2020). Further, shifts in patterns of brain activity from one brain “state” to another—when watching a movie, for example—tend to correspond well with event boundaries identified behaviorally (Baldassano et al., 2017; Geerligs et al., 2021; Lee et al., 2020). Event boundaries specific to each region are evident in these brain state shifts, with more frequent shifts in lower-level sensory regions and less frequent shifts in higher-level cortices, indicating that higher level regions process events on longer timescales.

Event Memory

When it comes to remembering events, there is support for the idea that recall is structured around the same event boundaries identified during the initial processing of the episodic experience (reviewed in Zacks et al., 2020). Additionally, participants who are “better” at segmenting—meaning their demarcation of event boundaries is more consistent with the group average—tend to have better subsequent memory for the events they segmented (Sargent et al., 2013). On top of the long term memory implications, there is extensive research suggesting a role for working memory capacity in event segmentation and long term recall as well (reviewed in Kurby & Zacks, 2008). This role

is proposed because of the idea that one maintains one's current event model in working memory until a boundary allows for this information to be offloaded into long term memory (reviewed in Kurby & Zacks, 2008).

Part of what often leads to the delineation between one event and another and the subsequent updating of the event model, argue some researchers, is the presence of prediction error at the boundary—when prediction error is high, people are more likely to demarcate the end of an event, as the next likely outcome is uncertain (Zacks, 2020). Others point out, however, that prediction error may be an inaccurate or maladaptive way of creating event boundaries when the environment is uncertain and prediction error is always in an elevated state (reviewed in Shin & DuBrow, 2021). These researchers argue that change alone—whether spatial, temporal, or character- or action-based—is sufficient to create a boundary, whether or not the change is expected (Shin & DuBrow, 2021). This helps explain why we can segment experience into events even when the experience follows a script. In this same vein, one paradox in the literature is that events tend to be better remembered when they deviate strongly from the expected event structure, but also when they follow a common script (reviewed in Zacks, 2020). In other words, the most atypical and the most prototypical events are best remembered.

Event boundaries are behaviorally salient and are identified relatively consistently across adult participants, and this is reflected in neural activity as well. Neuroimaging work has shown that hippocampal activity increases at event boundaries (Baldassano et al., 2017), suggesting that the previous event is being pushed into long term memory as the brain offloads information from working memory and shifts its current processing onto the next event. Details that occur near event boundaries also tend to be better

remembered than those from the middle of events (reviewed in Kurby & Zacks, 2008 and Zacks, 2020), suggesting that boundaries are particularly salient to us and it is advantageous to recall the occurrences close in proximity to them. Additionally, boundaries tend to yield “chunking” of events, such that elements occurring between two boundaries are grouped together—within-event relationships are better retained than those across event boundaries (reviewed in Shin & DuBrow, 2021). Participants are also better able to remember the temporal order of items and more likely to rate them as having occurred in closer temporal proximity when the items are drawn from the same event versus from a different event, controlling for actual temporal distance between the items (Ezzyat & Davachi, 2014). There is therefore reason to believe that event segmentation during ongoing perception is beneficial to later episodic memory.

Development of Event Cognition and Memory

Spontaneous event segmentation appears to develop early in life, with even prelinguistic children showing good event understanding (Keven, 2016), and neural activity of infants tracking ongoing events (Yates, Skalaban, Ellis, Bracher, & Turk-Browne, pre-print). Children are able to behaviorally segment events while watching a movie, and do so similarly to adults, although their segmentation is more variable (Zheng, Zacks, & Markson, 2020). Similarly, some aspects of children’s episodic memory—such as item recognition—are robust early on, while performance on more complex tasks—like item-context binding or temporal order memory—is markedly inferior to that of adults in early childhood, especially before age 6 or 7 (Benear, Ngo, Olson, & Newcombe, 2021; Blankenship, Calkins, & Bell, 2022; Newcombe, Lloyd, & Ratliff, 2007; Ngo, Newcombe, & Olson, 2018). This is generally believed to be a

consequence of protracted hippocampal maturation across this age range (Daugherty, Flinn, & Ofen, 2017; Keresztes et al., 2017; Riggins et al., 2018). It is possible that a still-developing ability to understand and segment events might also play a role in children's improvement in episodic memory in early childhood. Indeed, although children can often correctly identify details from events when prompted, their free recall of those same events is poor (Zheng et al., 2020), perhaps reflecting a lack of cohesive narrative structure, which could not be generated due to a paucity of robust event memories. Content pulled from within events is better recalled than that from across events for children as well as adults, but reaction time in children is affected less than adults' by event boundaries (Ren, Wharton-Shukster, Bauer, Duncan, & Finn, 2021), suggesting that boundaries might not structure children's memories as strongly as they do adults' memories—or, perhaps, what adults define as a boundary might not be perceived as such by a child. In addition, the ability to place events in the correct temporal order increases across early childhood and into adulthood (Bettencourt, Everett, Chen, & Pathman, 2021; Blankenship et al., 2022; Price & Goodman, 1990), and more details are included in children's descriptions of events with increasing age (Price & Goodman, 1990).

Although the ability to recall novel experiences is still developing in childhood, even very young children have robust and orderly scripts for everyday events (Keven, 2016). This aligns with evidence that children develop the ability to accrue semantic knowledge, such as vocabulary and scripts, in the first two years of life (Bloom, 2002; Carey, 1982), much earlier than they can effectively recall unique autobiographical episodes. Interestingly, although scripts seem to scaffold mnemonic performance for children, children are better able to remember events that deviate from the script in some

way (Farrar & Goodman, 1992; Hudson, Fivush, & Kuebli, 1992), emphasizing that atypicality increases memorability for children like it does for adults, but perhaps only within the confines of an otherwise familiar script.

Motivation for the Current Study

Children are able to segment events like adults are, but with more variability in their boundaries (Zheng, Zacks, & Markson, 2020), and children are able to remember events like adults are, but with less consistency and less detail (Ngo, Lin, Newcombe, & Olson, 2019), but are these events mapped in children's brains the same way as they are in adults' brains? The assertion that adults segment events spontaneously without instruction to do so is bolstered by neuroimaging evidence (Baldassano et al., 2017; Geerligs et al., 2021; Lee, Aly, & Baldassano, 2020), but studies investigating neural signatures of event segmentation in children are limited. While nascent evidence suggests that children's neural event signatures show greater group agreement with increasing age and that behavioral boundaries align with neural boundaries in children as young as 5 (Cohen & Baldassano, 2021), more research is needed to elicit whether children perceive event boundaries in fundamentally different ways from adults, and whether this is reflected in their neural activity. Further, although some data suggests children whose event segmentation is more similar to that of adults have better memory of the events (Zheng, Zacks, & Markson, 2020), whether children's event segmentation ability predicts their subsequent memory for events when tested using methods of differing difficulty, and when accounting for age and working memory capacity, is a question that merits further study.

With this study, our first goal was to determine whether neural activation patterns reflect behaviorally-segmented event boundaries in children. We hypothesized that children would show some agreement with adults for their event boundary selection, but that this agreement would be lower than either children's or adults' intragroup agreement, based on prior work showing similar results when child and adult participants segmented short films (Zheng et al., 2020). We also predicted that the neural activity from one sample of children would reflect the behaviorally-delineated boundaries from another group of children. Specifically, we expected that neural activation patterns would be more similar within events than across event boundaries, and that this pattern would be present for boundaries delineated by children to a greater degree than those delineated by adults.

Our second goal was to evaluate whether event segmentation agreement predicts subsequent memory for event details. We hypothesized that adults whose event segmentation is more closely aligned with their group norm, and children whose segmentation is most aligned with the adult norm, would show better memory performance for details of the events than those whose segmentation is more variable. We expected this effect because participants who are better able to track ongoing events in a consistent way may be better able to recall these events as a cohesive narrative. Because of the salience of event boundaries in both neural data and behavioral measures in prior work, we predicted that both children and adults would remember details from the show that occurred near event boundaries better than those that occurred during the middle of an event.

We probed these questions using an existing dataset of naturalistic fMRI data collected in young children, as well as new behavioral data collected in both children and adults. Both adult and child participants in our behavioral study watched the same animated television show that the child participants in the imaging study watched. The behavioral participants also segmented the show into events, completed several tasks to probe their memory for the show, as well as completing neuropsychological tests of working memory and verbal skills. Our goal was to use this rich data to examine the association of children's event cognition and memory with adults', as well as to determine whether children's behaviorally delineated event boundaries can be mapped onto neural data from a separate cohort of children to identify event boundaries in their brain activity.

CHAPTER 2

METHODS

Participants

Behavioral data was collected from young adults ($N = 42$; 24F, 17M, 1NB; $M_{\text{age}} = 19.29$) recruited from Temple University's SONA subject pool, and children ages 4-5 ($n = 23$; 13F, 10M; $M_{\text{age}} = 4.87$) and ages 6-7 ($n = 21$; 11F, 10M; $M_{\text{age}} = 7.16$) recruited from the Temple Infant and Child Lab database, as well as via tables hosted at local farmer's markets and community events, fliers posted in public spaces such as libraries, emails to previous participants with permission to recontact, posts on social media, etc. All child participants were free from developmental and neurological disorders as reported by a parent and all participants had normal or corrected-to-normal vision and hearing. Adults and parents of children all reported that they had not watched the show *All Hail King Julien (AHKJ)* prior to the study. Several children mentioned familiarity with the show initially but, when probed, revealed that their familiarity was with characters who had also appeared in the motion picture *Madagascar*, which had different plotlines and only a few characters that overlapped with *AHKJ*.

The imaging data analyzed in this project was collected at the University of Pittsburgh in 2018 and 2019 and included 46 children with complete anatomical and functional MRI data ($M_{\text{age}} = 5.99$; 23F, 23M; 14 4-year-olds, 9 5-year-olds, 13 6-year-olds, 9 7-year-olds, 1 8-year-old). All participants' data was visually inspected, and two participants were removed after inspection for significant data quality issues due to motion (both F, both age 4). Participants for whom greater than 25% of TRs were motion

outliers—defined as TRs with a framewise displacement greater than 0.5mm—were also removed from the dataset ($n = 7$, $M_{\text{age}}=5.96$; 4F, 3M; 2 4-year-olds, 1 5-year-old, 3 6-year-olds, 1 7-year-old). Last, one participant was removed from the dataset due to significant dropout in signal from the hippocampus (M, age 5). This left a remaining dataset of 36 participants ($M_{\text{age}}= 6.12$; 17F, 20M; 10 4-year-olds, 7 5-year-olds, 10 6-year-olds, 8 7-year-olds, 1 8-year-old) to be used in the imaging analyses. Thus, there were roughly equal numbers of younger children (ages 4-5) and older children (ages 6-7) in the dataset, with one 8-year-old also included in the older child group.

Stimuli

Event Segmentation Task

The Netflix television series *All Hail King Julien (AHKJ)*, Season 1, Episodes 1 and 2 were used for training and the event segmentation task, respectively. In the existing neuroimaging dataset, participants trained in the mock scanner using Episode 1 and viewed Episode 2 during their functional MRI scan. As such, Episode 2 is the episode of interest for demarcating event boundaries. The Coder from the software program PsychoPy (Peirce et al., 2019) was used to create the task that collected button presses from both adult and child participants when demarcating event boundaries while viewing this episode, with the button presses time-locked to the progression of the show. Response times were collected as output in seconds, out of 1283 total seconds across the entire episode. Prior to demarcating boundaries during Episode 2 of *AHKJ*, participants practiced this task using the last five minutes of Episode 1. In addition, before this

practice round, child participants had another opportunity to practice pressing the button to demarcate events using two clips from *Pingu*, a British animated television series.

Episodic Memory Tasks

We implemented three tasks to index episodic memory. The first was a free recall task, in which participants were prompted to say aloud everything they remembered from the second episode of *AHKJ* and their responses were audio recorded. This data was not analyzed for the purpose of this project. Next, participants completed a forced-choice recognition task and a temporal order memory task. The PsychoPy Builder (Peirce et al., 2019) was used to create and present both tasks and collect participant responses. The forced-choice recognition task asked participants 20 three-alternative multiple-choice questions about events from the *AHKJ* episode 2, relatively evenly spaced across events from the episode, asking participants to correctly identify event details, characters, and scenes. The temporal order task was comprised of 40 two-alternative multiple-choice questions, presenting screen shots from *AHKJ* episode 2 with captions. Participants were asked to determine which of the events depicted in each screen shot happened first in the episode.

Working Memory and Verbal Intelligence Control Measures

Working memory is suggested to be a strong correlate of performance on event segmentation tasks (reviewed in Kurby & Zacks, 2008), as being able to determine boundaries in continually shifting incoming perceptual information requires manipulation of information in working memory. We used the Cambridge Neuropsychological Test Automated Battery (CANTAB) Spatial Working Memory (SWM) subtest to assess working memory in both children and adults. The task takes less than five minutes to

complete, relies on spatial rather than verbal stimuli—so children who can't yet read can still complete it independently—and increases in difficulty as the task progresses. This test has previously been used successfully in studies with children in the same age range as those in our sample (Luciana & Nelson, 1998; Moore et al., 2021). The CANTAB SWM served not only as a useful control measure, but also as a measure of interest, helping to delineate whether better event segmentation ability with increasing age is specific to that ability, or whether it can simply be attributed to general cognitive development, which working memory performance would help to elicit.

Additionally, to account for effects of general intelligence, we collected the Verbal and Riddles subtests of the Kaufman Brief Intelligence Test, Second Edition (KBIT-II). This measure is intended to capture verbal IQ and is designed for use with both children and adults; it has previously been used in studies with children in the same age range as those in our sample (Benear et al., 2021; Ngo, Newcombe, & Olson, 2019). Scores on the KBIT-II are standardized to account for age.

Behavioral Procedure

Encoding and Control Tasks

Participants were shown a slideshow in PowerPoint to introduce the characters and explain the plot of Episode 1 of *AHKJ*. Then, participants watched the last five minutes of *AHKJ* episode 1, to familiarize them with the portion of the show that would later be used as practice for the event segmentation task. Next, participants watched the entirety of episode 2 of *AHKJ*. Participants watched the show on a desktop or laptop computer while wearing over-ear headphones.

After watching episode 2, participants completed the working memory task (CANTAB SWM), as well as the verbal intelligence task (KBIT-II), which inserted a natural delay of roughly 20 minutes between viewing the episode and being tested on its content. After this, participants were offered a break of up to 15 minutes, bringing the total time to 20-35 minutes between watching the show and testing memory. Children were allowed to have a snack or drink, use the restroom, play with toys, and/or talk to their parents about things not related to *AHKJ*. Young adults were allowed to use the restroom, get a drink of water, or have a snack. Most adults did not elect to take a break, with the longest adult break being roughly five minutes; all children took a break, although their breaks varied from three to 15 minutes. The break was included because of the lengthy nature of our study—1.5 to 2.5 hours depending on the participant’s age. The break allowed children in particular to move around and rest their minds so they could return to the study with focus.

Mnemonic Tasks

After the break, participants completed the mnemonic tasks, beginning with free recall. This task was not analyzed for the purposes of this project. After free recall, participants completed the forced-choice recognition task. Participants were asked 20 questions about the episode in a three-alternative forced-choice format, (e.g. “Which of these characters can read minds and tell the future?”), followed by images of three characters from the show; “What is the first way King Julien tried to get all the lemurs to like him again?”), followed by three possible written answers to this question. Ten of the questions had images as the answer choices and 10 had text. For those with images as the answer choices, they were a target and two lures (depicting characters, items, or scenes

from the show), and for those with text as the answer choices, they were a target and two foils (with the two foils describing incorrect details or events that didn't happen in the show). Adult participants completed this assessment independently, clicking keys 'a', 'b', or 'c' to select their answer choice. The researcher read the questions to child participants, and the children pointed to their answer or said it aloud. The researcher then selected the letter corresponding to their choice. Questions with text as the answer choices were read to children at least twice.

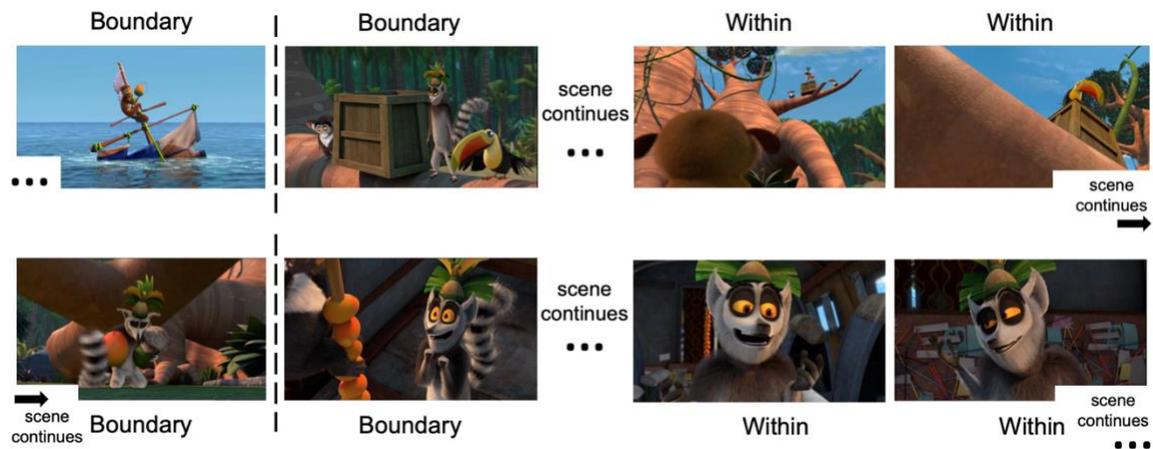


Figure 1. Across-boundary vs. within-event pairs. All pairs of events for the temporal order task were pulled from either side of an event boundary, or from within the middle of an event in the show. These boundaries were pre-determined by a group of graduate student raters. This schematic also depicts details occurring in the show on either side of two example boundaries and pseudo-boundaries (see Methods: Imaging Data; and Figure 6).

Next, participants completed the temporal order task. There were four counterbalanced versions of the task, each with 40 questions. In this task, screen shots pulled from *AHKJ* episode 2 were depicted on screen in pairs, each with a short caption (e.g. “King Julien enters Hector's house and sees a tower of matches”, “Gigi realizes the tree is about to fall down”) and participants were asked to choose which mini event depicted in the screen shots happened first. These mini events were selected from either

within an event or adjacent to an event boundary (Figure 1). The event boundaries were nine boundaries that were chosen by at least three of four graduate student raters who viewed the show and denoted where they believed event boundaries to be located, plus the beginning and the end of the episode. For the comparisons of screen shots, 13 were boundary to boundary comparisons, 13 were within to within comparisons, and 14 were boundary to within comparisons. Our expectation was that boundary details would be recalled better than details from within events, so we wanted to structure the task to evaluate this. Adults completed the task independently, selecting the ‘1’ or ‘2’ key to choose which screen shot they believed depicted the detail that happened first in the show. The researcher read the captions aloud to child participants, pointing to each screen shot corresponding to the caption as they read. Children pointed to the screen shot they believed came first or said their response aloud, and the researcher selected the key corresponding to the child’s choice.

Event Segmentation Task

After completing the mnemonic tasks, participants were asked to segment *AHKJ* episode 2. First, participants were trained on how to complete this task. The researcher told the participant that we were interested in how they think about the events from *AHKJ*. A realistic series of three events that would be part of a generic morning routine were described to both children and adults (e.g. pick out your clothes and get dressed, go to the bathroom to brush your teeth, go to the kitchen for breakfast). Child participants were asked to think of one more event that might happen before leaving the house in the morning to probe their understanding and were corrected if the chosen event was too broad or too narrow. We told all participants that for this task, we were curious where

they thought the boundaries in the show were, and that there was no right or wrong answer.

For child participants, there was next an example task using two clips from a British animated television series called *Pingu*, which were presented without audio. During the first clip, the researcher segmented the clip into events by pressing the space bar on the keyboard, narrating why they made their choices as they went along (e.g., “The penguins have pulled the pan off the stove and are sitting down to eat the popcorn, so I’ll mark that as a new event.”). Some children quickly understood the task and were eager to take over and press the button themselves. If they did this for at least two events during the first clip, the example task was terminated and they moved on. For children who watched the experimenter during the first clip but didn’t press the button themselves, they were given a second *Pingu* clip to segment by pressing the space bar, with the experimenter encouraging them along the way, talking them through the events, and asking questions to ensure they understood what we were requesting of them with the task.

After the example practice with *Pingu* for children, and after the realistic series of event description for adults, all participants moved on to segment *AHKJ*. There was a practice round using episode 1, to give adults a chance to ask questions, and familiarize children with segmenting the events of *AHKJ*. Children were prompted to segment events if they did not seem to be pressing the button; all adults understood the task and did not need prompting. After the practice round, participants were asked if they had any questions. If not, they were told to proceed and advance through the task on their own. For child participants, we asked them to tell us when they were supposed to press the

button. Answers approximating correctness were accepted (e.g. “when something new happens in the show”), and for children who did not answer, we gave multiple choices aloud and asked them to tell us which was correct. If they could not, we reminded them again of the task at hand.

Participants then moved on to segment the entirety of *AHKJ* episode 2, adults on their own and children with the researcher nearby in the room. The movie was split into five segments of 3.5-5 minutes each, totaling to 21 minutes and 23 seconds—this was not the length of the full episode, but instead excluded the last roughly one minute of the show to match the stimuli the Perlman lab used when collecting their imaging data. Four counterbalanced versions of the event segmentation task were used, with the breaks occurring at slightly different times to account for any incidental effects of breaks being placed close to boundaries in the episode. Each break consisted of a gray screen with text that said, “Time for a break!” and an image of King Julien. Following each break, the previous 10 seconds of the episode from before the break were shown to remind participants of what had been happening. Adults were told they could stop to rest briefly during the breaks, or they continue without breaking. Children were asked at each break how they were doing and were allowed to continue if they wanted. For those that wanted a break, they could have a drink of water or stand up to move around. They were not encouraged to leave the room, but were permitted to leave to ask their parent something or to use the restroom if they requested it. The breaks were deemed necessary, as it is exceedingly difficult to remain in a particular task set for 20 minutes when passively viewing stimuli. These breaks encouraged children in particular to remain attentive and respond to the movie with button presses at event boundaries. Because of the length of

the full study and complexity of this particular task, some of the younger children in the study grew restless, but only two children (1F, 1M; $M_{age} = 4.5$) failed to complete the task due to fatigue and/or inattentiveness. Thus, the majority of our participants showed good understanding and persistence through the duration of the task.

Imaging Data Acquisition and Preprocessing

The MRI data was acquired on a Siemens Magnetom Prisma Fit 3 Tesla scanner with a 64-channel head coil at the University of Pittsburgh. Structural data was acquired using a T1-weighted protocol (TR = 1900 ms; TE = 3.58ms; 1 mm isotropic voxels; FOV = 256 mm). Functional MRI data were acquired using a multiband, multi-echo planar imaging sequence (TR = 800 ms; TEs = 13.20 ms, 38.76 ms, 64.32 ms; flip angle = 49°; 40 slices, 3 mm isotropic voxels; FOV 204 mm). The functional scan was acquired while participants watched a 1283 s video clip (Season 1, Episode 2 of *All Hail King Julien*).

The imaging data was preprocessed using fmriprep 20.2.6 (Esteban et al., 2019), which is based on Nipype 1.7.0 (Gorgolewski et al., 2011). Because the functional data contained three echoes, a T2* map was estimated from the preprocessed BOLD by fitting the data to a monoexponential signal decay model with nonlinear regression, using T2*/S0 estimates from a log-linear regression fit as initial values. For each voxel, the maximal number of echoes with reliable signal in that voxel were used to fit the model. The T2* map was then used to optimally combine preprocessed BOLD across echoes following the method described in (Posse et al., 1999). The optimally combined time series was carried forward as the preprocessed BOLD. Since participants were ages 4-7 years old, the data was aligned to cohort 2 of the pediatric MNI template

(MNIPediatricAsym:cohort-2) through nonlinear registration, and analyses were carried out in this standard space. Following preprocessing in fmriprep, the data was spatially smoothed in FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) using a 6mm kernel.

Data Analysis

Behavioral Data

All behavioral data analyses were conducted using R. We collected data from 86 total participants (42 adults and 44 children), but two adults (both F, $M_{\text{age}}=18.5$), one older child (M, age 7), and two younger children (1M, 1F; $M_{\text{age}}=4.5$) did not complete the segmentation task and thus were not included in our final models. The two adults and older child were not able to complete the task due to technical issues; one younger child did not complete the task due to fatigue, and the other did not understand/comply with task instructions. There was also one child participant (F, age 5) for whom we did not collect KBIT data due to experimenter error, so they were excluded as well. Thus, our final sample for all of our models consisted of 40 adults (22F, 17M, 1NB; $M_{\text{age}}=19.32$) and 40 children (22F, 18M; $M_{\text{age}}=6.00$).

Our three dependent variables of interest were temporal order memory score, recognition memory score, and event segmentation agreement score. For temporal order memory, there were 40 total questions broken down into three categories. We calculated the number of correct responses to questions corresponding to within-within paired screen shots (13 questions), within-boundary paired screen shots (14 questions), and boundary-boundary-paired screen shots (13 questions) separately (see Figure 1). We also

calculated a composite score, which was the total number of correct responses to all 40 questions. Chance performance was 50%, so children answering less than 20 questions correctly were considered to perform below chance. For recognition memory, we calculated the total number of correct responses out of 20 questions. Chance performance on this task was 33%, so children answering fewer than seven questions correctly would be considered to have performed below chance.

For event segmentation agreement, we calculated three measures of agreement – adult to the adult norm (adult within), child to the child norm (child within), and child to the adult norm (child to adult). We calculated this score using a method known as the agreement index, implemented in several prior publications (e.g. Sasmita & Swallow, 2022; Zheng et al., 2020), with the help of code provided by Yinyuan (Sean) Zheng. Since the movie was broken into five clips for our task, first the response time in seconds for each clip had to be adjusted to account for the distance from the beginning of the movie rather than each individual clip. Once a list of values was generated for each participant that indicated where they placed all their boundaries in seconds across the movie, we then vectorized this list by binning each second of the 1283s movie into 1s bins, and giving each participant a value of 1 if they indicated a boundary in that bin, and 0 if they did not. Thus, each participant would have a vector of equal length, with binary indicators across the vector for the location of their boundary demarcations.

We calculated a norm for each group by averaging all runs across all individuals within a group (adults or children). Individual vectors were then compared to the group norm by computing the point-biserial correlation between an individual's segmentation vector and the group norm, with the resulting value being between 0 and 1, and higher

numbers indicating greater agreement. This was calculated within group for both adults and children, and the children's individual vectors were compared to the adult group norm as well, to obtain a measure of "adult-like" segmentation. The correlations were rescaled based on individual differences in the number of events identified (see Zheng et al., 2020 for additional details). When creating visualizations and estimated proportion of participants from each group indicating a boundary at a given location, we acknowledged that small deviations in boundary locations (on the order of 1-2s) might simply be a lack of precision and might actually indicate the same event being demarcated by different participants, especially for children. Thus, we "smoothed" participants' vectors by extending the binary indicators of boundaries out by two bins from each boundary in either direction, creating a five-second window for each boundary (Figure 2). This allowed for us to account for participants who might have indicated a boundary with low precision, and to more accurately represent the proportion of participants agreeing on a given event.

Our control measures were the KBIT-2 Verbal sections and the CANTAB Spatial Working Memory subtest. The KBIT sections were scored according to the manual, and then standardized to account for age differences, thus serving as a measure of verbal IQ that is agnostic to developmental age. For the CANTAB, we selected the SWMTE12 metric as our metric of interest, which is a value representing the number of errors made on the 12-box trials. The CANTAB progressively increases in difficulty across the task, with increasing numbers of boxes on each trial that one must maintain in working memory to correctly select only the boxes that have not yet been selected. The 12-box trial is the final and most challenging trial, and is the one with the most variability across

both child and adult participants, so it was selected as the most sensitive metric of working memory.

We set up three linear models for each of our three dependent variables of interest: temporal order memory, recognition memory, and event segmentation agreement score. For each dependent variable, we had one model run only on the adult data, one run on only the child data, and one run on both the adult and child data. The independent variables included in all models were age, KBIT score, and CANTAB score. For the adult models predicting temporal order and recognition, within-group segmentation agreement was included as a predictor, whereas for children, child-to-adult agreement was included instead. In the model including both adults and children, we included a metric we deemed “adult-like” agreement, which for adults was the value for within-group agreement and for children was the value for child-to-adult agreement. For the models predicting segmentation agreement, we used within-group agreement as the outcome of interest for the adult model, child-to-adult agreement for the child model, and adult-like agreement for the models including both adults and children. All models predicting segmentation agreement included both temporal order and recognition memory as predictors in addition to those previously noted.

Imaging Data

After the imaging data was preprocessed using fmripred and smoothed in FSL, we conducted our first level analysis in FSL using FEAT. The eventual goal was to conduct a representational similarity analysis (RSA) comparing neural patterns of activity during windows of time on either side of a boundary vs. adjacent to one another within an event (Figure 1). To do this, we used the boundaries generated by our adult and child

behavioral participants to define the time windows we would use in our RSA (Figure 6). We generated the chosen boundaries separately for each group, predicting that the boundaries indicated by children might better represent event boundaries in our neural dataset of children than boundaries indicated by adults would.

We ran two separate first-level analyses with FEAT—one based on the boundaries behaviorally delineated by adults, and one based on the boundaries behaviorally delineated by children. For both adults and children, we used the smoothed event segmentation data to generate proportions of participants indicating a boundary at a given time point, after which we could rank the boundaries by how frequently they were selected by either adults or children. For both adult- and child-delineated boundaries, we selected the top ten most commonly chosen boundaries (as time points in seconds, e.g. 600s) as those that would be used in our RSA. The boundaries could not occur too close to one another or they would overlap when comparing neural pattern similarity in different time windows across the time series. Therefore, if a boundary was less than 32 seconds from another boundary (see below for rationale for this value), it was excluded, and the next most-frequently selected boundary was chosen instead. After selecting these boundaries for both groups, we used them to generate our time windows.

A given time window was defined as a five-second time period ending three seconds before a boundary or beginning three seconds after it (see Figure 6). We included the time cushion on either side of the boundary to account for differences in the exact boundary location as determined by different participants. For each of our within-event time windows, we selected a pseudo-boundary occurring halfway between the previous boundary and subsequent boundary. We again chose five-second windows ending three

seconds before the pseudo-boundary or beginning three seconds after it. For each pair of boundaries, there was a minimum of $(3+5) * 4 = 32$ s required to exist between them in order to include two within-event windows without any overlap. In the end, we had 10 sets of boundary windows (20 total) and 10 sets of within windows (20 total). A text file was generated for each time window and included as an EV in FEAT for the first-level analysis. From the confounds file created by fmriprep during preprocessing, we pulled the six motion parameters, white matter and CSF, as well as cosine-based regressors, which effectively high-pass filter the data at 128s when included as confounds in the first-level analysis (Esteban et al., 2019). We added all 40 EVs and the confounds to the first-level analysis in FSL, which generated 40 t-statistics, one for each time window.

Next, we defined six regions of interest: posterior cingulate cortex (PCC), left and right angular gyrus (LAG, RAG), left and right hippocampus (LHC, RHC), and early visual cortex (EVC). All regions were defined using masks from the Harvard-Oxford cortical and subcortical atlases in FSLeyes, thresholded at 50 percent, and binarized. Then, FLIRT was used to convert them into the same pediatric MNI space to which our data was registered. PCC, L/RAG, and EVC were selected as regions commonly implicated in event processing (e.g. Baldassano et al., 2017; Ben-Yakov & Henson, 2018; Geerligs et al., 2021), while L/RHC was selected for its role in memory—of particular interest to our research questions—and because of prior work showing increased activation of the hippocampus at boundaries (Ben-Yakov & Henson, 2018), suggesting it may also track them. There was signal dropout in some participants' data, so we multiplied each ROI mask by each participant's fMRI data to obtain a modified mask in which all voxels would map onto data for all participants. PCC and LHC did not need to

be modified. LAG, RAG, and RHC had 0.2%, 1.8% and 4.6% of voxels removed, respectively. The only ROI with significant removal was EVC, in which 15% of the original ROI's voxels were removed, as almost all participants had empty voxels in this mask where it extended beyond the back of the brain. One participant was removed from the dataset because of significant dropout in LHC that would have resulted in a large portion of the ROI being removed for all participants if this participant was retained (see Methods: Participants).

We ran our RSA separately for each ROI, and separately for the adult-delineated boundaries and the child-delineated boundaries. For each ROI for each set of boundaries, we first calculated the distance between every pairwise combination of t-statistics using the `pairwise_distances` function from the `scikit-learn` package in python (Pedregosa et al., 2011). This generates a dissimilarity matrix, which represents a dissimilarity score for the spatial correlation between every pair of t-statistics across the timeseries, on a scale of 0 to 2, with 0 being exactly the same and 2 being perfectly dissimilar (Figure 6). From this matrix, we pulled only the values corresponding to the pairs of time windows taken from either side of the same boundary or same pseudo-boundary (Figure 6). Thus, for each participant for each ROI for each boundary type (child-generated or adult-generated), we had 10 within-event similarity values and 10 across-boundary dissimilarity values. These could be used to generate a comparison for each ROI to determine whether pattern similarity was different when comparing within-event to across-boundary brain activity.

Finally, we also ran a Hidden Markov Model (HMM) based on a workshop designed by Chris Baldassano (https://naturalistic-data.org/content/Event_Segmentation.html). HMMs (described in Baldassano et al., 2017; Lee, Aly, & Baldassano, 2021), rely

on two assumptions: “(1) while processing narrative stimuli, observers experience a sequence of discrete events, and (2) each event has a distinct neural signature,” (Baldassano et al., 2017). Essentially, the model detects changes in “brain states” across the time series, with the point at which a given state changes to another state corresponding to a boundary. You feed the model the number of events you want it to find, and the output is a set of boundaries that divides the timeseries into events. Additionally, you can feed the model several numbers of events and determine the optimal number of events for a given ROI using the loglikelihood to determine which number of events is the best fit to the timeseries for that ROI. Finally, you can compare the boundaries delineated by the HMM to boundaries delineated by human participants and determine whether they are significantly related to one another. In this way, it can be determined through data-driven methods that the brain’s neural response to changes in events correspond to the places in which participants behaviorally determine that one event has ended and a new one has begun.

CHAPTER 3

RESULTS

Event Segmentation Agreement

Qualitative Data Exploration

The event segmentation data for adults and children can be explored in many ways. First, after creating the binned vectors of event boundary demarcations for each participant, we calculated averages for each 1s bin for each group to determine the proportion of participants indicating a boundary at a given timepoint. We then realized that much of the agreement between participants was being lost to the inherent noise in the locations of boundaries—differences of one or two seconds in either direction. Thus, we “smoothed” the bins 2s in either direction from each participant’s boundary markers to allow for agreement with other participants who might have been indicating the same boundary but whose response time to the end of one event and start of another might have been slightly different (described in Methods, Data Analysis, Behavioral Data). We then plotted this smoothed timeseries showing the proportion of participants indicating a boundary at a given location (Figure 2).

This method yielded a higher proportion of participants agreeing on boundaries for both groups than the unsmoothed version, more accurately depicting agreement among participants. Upon visual inspection, it is obvious that children’s boundary demarcations are more variable than those of adults (Figure 2), which we verify in the following sections by calculating agreement scores. However, there is still substantial agreement among children at several points, and there are several locations where both

adults and children have high agreement, indicating that both groups believed a boundary to be in that location (red arrows, Figure 2).

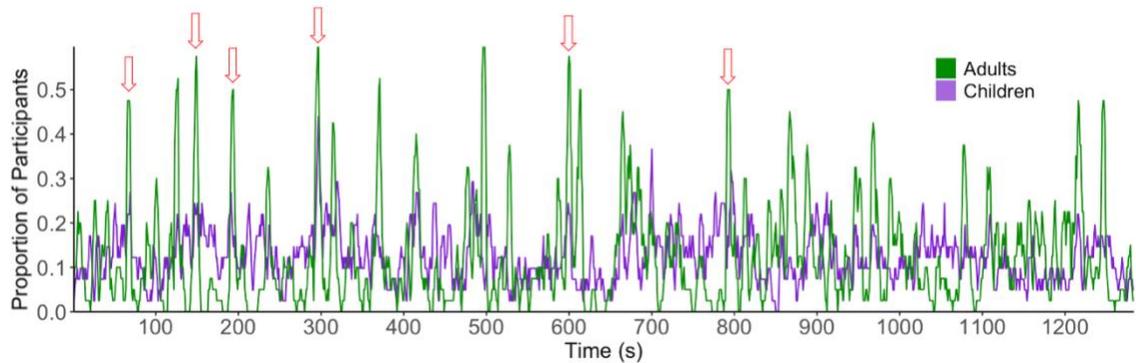


Figure 2. Plots of event segmentation across the entire television show. Proportion of participants indicating a boundary at a given location, visualized separately by group (adults in green and children in purple), and smoothed by adding a 2s cushion on either side of each boundary. Smoothing allowed for agreement between participants who indicated a boundary a few seconds apart from one another. Arrows indicate the points at which children and adults have similar peaks, indicating both groups agree there is a boundary in that location.

In addition to viewing this data in aggregate, we pulled a few individual children’s data from the dataset to view qualitatively and determine where children tended to place boundaries. Some children indicated roughly the same amount of boundaries as adults, roughly evenly spaced, while others indicated only a sparse handful of boundaries, and still others indicated a very large number of boundaries, often only seconds apart in succession. We investigated a few children’s data of each type. The pattern we observed was that children were largely practical in placing boundaries in locations where the average adult might, but they sometimes missed clear boundaries (e.g. a scene change) or placed boundaries in locations where there was clearly no change in event (e.g. in the middle of a conversation between two characters). Even children who had periods of rapidly pressing the button still had very reasonable boundary demarcations interspersed with these bursts, indicating that even when children might

have engaged with the task incorrectly for brief intervals, this didn't preclude their selection of sensible event boundaries. Children also often indicated a boundary at a point of salient or substantial action (e.g. a character suddenly flying across the screen) that was not necessarily an indicator of a new event. The overall qualitative assessment was that children, even as young as age 4, have a good understanding of what an event is and what constitutes an event boundary, but their ability to track the show consistently to accurately locate all boundaries is still developing.

Agreement Scores

To quantitatively calculate agreement among participants in a given group, or between a participant in one group with the average of another group, a segmentation agreement score can be calculated. We used the method commonly employed and described in the literature as the agreement index (Sasmita & Swallow, 2022; Zheng et al., 2020) to calculate agreement within and between groups (described in Methods, Data Analysis, Behavioral Data). Individual event segmentation profiles were compared to the group norm for adults or children, and a value ranging from 0 to 1 was the output. These values could be averaged for a given group to determine their mean agreement with a given group norm.

Our results indicate good within-group agreement for both adults ($M = 0.54$, $SD = .13$) and children ($M = 0.57$, $SD = .09$), but much lower agreement for children to the adult norm ($M = 0.16$, $SD = .06$), although any value above zero indicates that there is agreement (Figure 3A). This indicates that, although children's segmentation profiles look similar to one another's, they look different from adults'. However, there is still some agreement between children and adults, indicated by the non-zero agreement score.

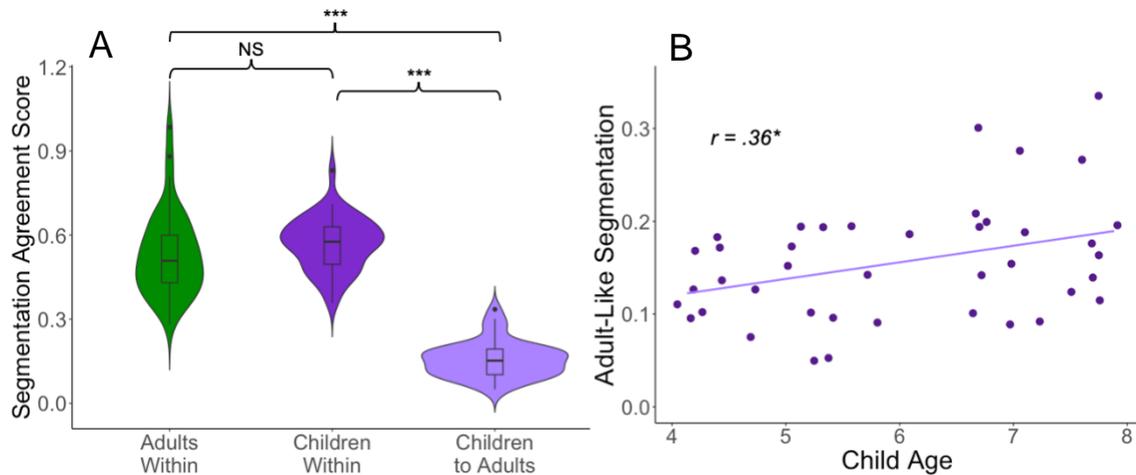


Figure 3. Event segmentation agreement. A. Agreement for adults to the adult norm (green, left), for children to the child norm (dark purple, middle) and for children to the adult norm (light purple, right). B. Scatterplot showing adult-like segmentation, i.e. children’s agreement with the adult norm, plotted against age for children, with age as continuous from 4.00-7.99. * $p < .05$, *** $p < .001$

Next, we wondered whether there might be a difference between younger children (ages 4-5) and older children (ages 6-7) in terms of their agreement with adults’ segmentation norm. Indeed, older children ($M = 0.18$, $SD = .07$) had significantly higher segmentation agreement with adults than younger children ($M = 0.13$, $SD = .05$; $t = 2.69$, $p = .01$). We further explored this through a correlation between children’s agreement scores and continuous age (from 4.00-7.99), and found that age is positively correlated with child-to-adult segmentation agreement ($r = .36$, $p = .02$; Figure 3B). This further supports the idea that as children develop across early childhood, their event segmentation profiles begin to look more like adults’.

Bootstrapping Analysis

To empirically test whether children’s segmentation is qualitatively different than adults’, we performed a bootstrapping analysis with 1000 replicates comparing the agreement segmentation vectors between individual pairs of participants, again using the

methods employed in a prior publication (Zheng et al., 2020) and with code provided by Yinyuan (Sean) Zheng. We used Cohen's Kappa as the metric of agreement since vectors of binary indicators are being compared, similar to the yes/no agreement between raters for which Cohen's Kappa is typically employed. Following from this, thinking of individual pairs of participants as "raters" between which one is looking for consistency in their ratings of a given variable, greater agreement on this metric indicates greater consistency in ratings, or in our case, consistency in the locations of event boundaries. After conducting this bootstrapping analysis using Cohen's Kappa, we calculated a difference score between the mean agreement between adult-adult pairs and the mean agreement between adult-child pairs. We calculated the same difference score for the mean of adult-child pairs minus child-adult pairs. If there was a significant difference between the within-group and across-group agreement, determined by a confidence interval not including zero, it would indicate that individuals within that group had more consistent segmentation with one another than with the other group.

We found that the difference score between adult-adult pairs and adult-child pairs was significant (95 percent CI:[.022, .057]). However, we found no significant difference between the child-child and child-adult pairs (95 percent CI:[-.005, .006]). This suggests that adults' segmentation is more consistent than children's segmentation, which is reflected in the visualization of both groups' segmentation data (Figure 2) but was not captured in the agreement metric (Figure 3). In sum, although children agree with one another and their segmentation shows some agreement with adults, it is much more variable than adults' segmentation, with smaller peaks in agreement and more noise between the peaks.

Full Behavioral Models

Temporal Order

Our first set of linear models was set up to examine the variables that contributed to temporal order memory performance. In a model including only adult data, with age, within-group segmentation agreement, verbal IQ, and spatial working memory (SWM) as predictors, and temporal order memory as the outcome, there were no significant predictors, although there was a trend toward significance for SWM ($p = .097$). In a model including only children, with age, child-to-adult segmentation agreement, verbal IQ, and SWM as the predictors, age ($b = 4.07$, $t(35) = 8.04$, $p < .001$) was the only significant predictor, with verbal IQ trending toward significance ($p = .056$). In a model including both adults and children, with age group (adults, older children, younger children), adult-like segmentation, verbal IQ, and SWM as the predictors, age: older kids vs. adults was significant ($b = -5.04$, $t(74) = 1.84$, $p = .008$) as well as age: younger kids vs. adults ($b = -13.75$, $t(74) = 2.02$, $p < .001$; Figure 4A). There was a trending relationship for verbal IQ ($p = .063$). We chose to model age as groups rather than continuously in the models including both adults and children in order to capture the developmental differences that occur in this age range in early childhood, as well as to avoid the issues with treating a variable as continuous when it has a large break disrupting the continuous nature of the data (in this case, no data for ages 8-17).

Because the full model had adults as the comparison group for our age factor variable, it did not allow for comparisons among all three groups. Thus, we further explored the difference between all three age groups by conducting a post-hoc test to examine the marginal means. This is simply an extension of the original linear model that

shows the direction of relationships in pairwise comparisons of all groups, so the statistics are identical for relationships already represented in the original model. For example, we found the same significant difference as in our full model for adults compared to older children ($b = 5.04$, $t(74) = 2.73$, $p = .008$), as well as younger children ($b = 13.75$, $t(74) = 6.80$, $p < .001$; Figure 4A). We were now, however, also able to compare older children to younger children, and we found that older children outperform younger children ($b = 8.72$, $t(74) = 7.06$, $p < .001$; Figure 4) as well. We also conducted a correlation between continuous age and temporal order memory performance in children, and found a significant positive correlation ($r = .84$, $p < .001$; Figure 4B).

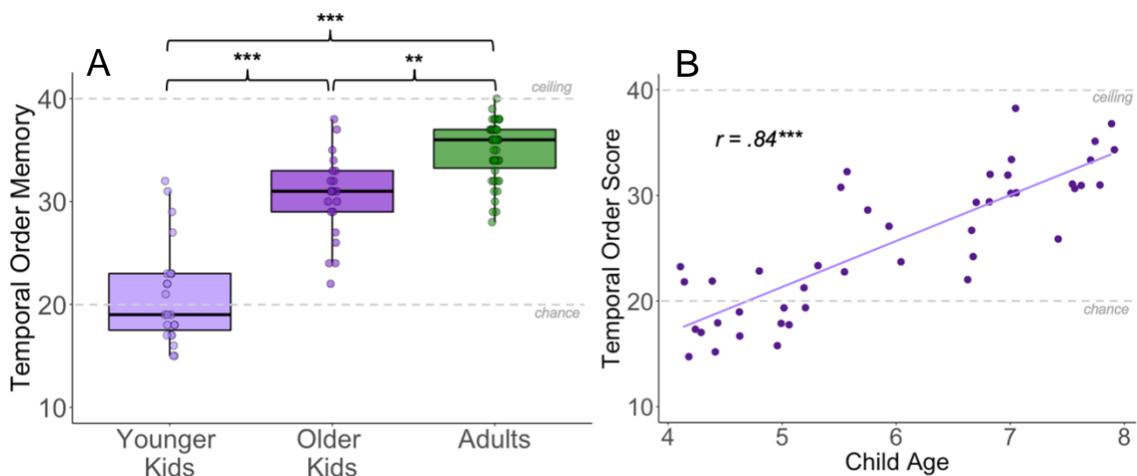


Figure 4. Temporal order memory. A. Temporal order memory performance for younger children (light purple, left), older children (dark purple, middle) and adults (green, right). B. Children’s temporal order memory scores plotted against continuous age (4.00-7.99). The lower horizontal dotted line in both plots visualizes chance performance (20/40), while the top line visualizes ceiling performance (40/40). ** $p < .01$, *** $p < .001$

We also noticed that many of the younger children were performing below chance level on this task (less than 20 questions correct), so we tested whether their performance was significantly different from chance. Indeed, we found that, although older children’s performance was significantly greater than chance performance ($t(19) = 10.95$, $p < .001$),

younger children were not performing significantly above chance on this task ($t(19) = 1.00, p = .17$). These results illustrate that there is a meaningful impact of age on temporal order memory performance, with significant improvements across early childhood and into young adulthood. While younger children are not able to perform this task reliably, older children are able to do so, although there is significant variability in performance among both older children and adults (Figure 4A).

Finally, because we hypothesized that memory would be better for details near event boundaries compared to those from within the middle of an event, we evaluated the sub-sections of the temporal order memory task separately. We ran an ANOVA including both child and adult data to determine whether performance was different for questions regarding boundary-boundary, boundary-within, and within-within screen shot comparisons. This analysis revealed that there was no difference in these subcomponents of the task ($F_{(2,240)} = 2.54, p = 0.081$), although there was a trend toward significance. This result suggests that our hypothesis was not supported—temporal order memory performance did not differ significantly by whether the details queried were pulled from near an event boundary or from within an event.

Recognition

Next, we aimed to investigate recognition memory performance. Our first model included only adult data, with age, within-group segmentation agreement, verbal IQ, and spatial working memory (SWM) as the independent variables and recognition as the dependent variable. Within-group segmentation agreement was a significant predictor of recognition memory ($b = -3.10, t(35) = -2.190, p = 0.035$) in adults, suggesting that adults who segment more consistently with one another have better memory for the

events they segmented. However, it's worth noting that many adults were at or near ceiling performance on this task (Figure 5A), so there is little variability in this particular outcome for adults. In the model including only children's data, we again included age, child-to-adult agreement, verbal IQ, and SWM as the predictors. Age was a significant predictor ($b = 2.76$, $t(35) = 10.269$, $p < .001$) of recognition memory, as was child-to-adult segmentation agreement ($b = -10.76$, $t(35) = 2.061$, $p = 0.047$). There was also a trend toward significance for verbal IQ ($p = .057$). This outcome suggests that children who segment more like adults have better memory for the events they segmented.

Finally, we again ran a model that included both the adults' and children's data. In this model, age: older kids vs. adults was significant ($b = -3.40$, $t(74) = 3.61$, $p < .001$) as was age: younger kids vs. adults ($b = -9.60$, $t(74) = 9.28$, $p < .001$). Additionally, adult-like segmentation ($b = -4.44$, $t(74) = 2.23$, $p = .029$) and verbal IQ ($b = 0.06$, $t(74) = 2.72$, $p = .008$) were also significant predictors of recognition memory. These results suggest that performance on recognition memory improves with age, and that segmenting in a manner more similar to the adult norm and having a higher verbal IQ both relate to better recognition as well.

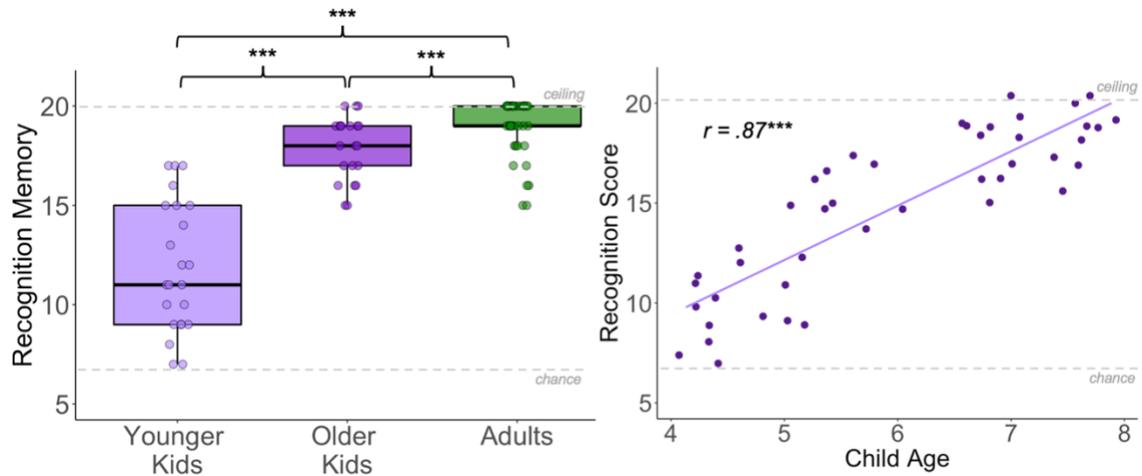


Figure 5. Recognition memory. A. Recognition memory performance for younger children (light purple, left), older children (dark purple, middle) and adults (green, right). B. Children’s recognition memory scores plotted against continuous age (4.00-7.99). The lower horizontal dotted line in both plots visualizes chance performance (6.67/20), while the top line visualizes ceiling performance (20/20). *** $p < .001$

As with temporal order memory, we again examined the differences among all three groups via a post-hoc examination of the marginal means. The findings for adults versus older and younger children were again statistically identical to the results in the full model, but with directionality showing us that adults outperformed older children ($b = -3.40$, $t(74) = 3.61$, $p < .001$) as well as younger children ($b = -9.60$, $t(74) = 9.28$, $p < .001$). This analysis also revealed that older children outperformed younger children as well ($b = -6.20$, $t(74) = 9.81$, $p < .001$; Figure 5A). Last, we conducted a correlation between continuous age and recognition memory performance in children, and found a significant positive correlation ($r = .87$, $p < .001$; Figure 5B). Both older children ($t = 29.52$, $p < .001$) and younger children ($t = 6.60$, $p < .001$) performed statistically above chance on this task—in fact, all children in the dataset answered seven or more questions correctly, putting them above chance performance (Figure 5). These results altogether suggest that even very young children can successfully perform recognition memory

tasks, but that performance continues to increase across early childhood and up to young adulthood. Additionally, those who segment events in a more “adult-like” manner seem to perform better on this task, suggesting that understanding of event structure may support recognition memory.

Segmentation Agreement

For our final behavioral models, we asked whether performance on either memory measure was related to adult-like segmentation. Our first model containing only the adult data had temporal order memory, recognition memory, age, verbal IQ and SWM as predictors of within-group segmentation agreement. Only recognition memory was a significant predictor of agreement ($b = -0.035$, $t(73) = 2.15$, $p = .039$), capturing the inverse of the relationship we found in the above model in which within-group segmentation agreement predicted recognition memory in adults. In our model containing only children’s data with the same predictors as in the adult model and the outcome variable being child-to-adult segmentation agreement, we found that only age was a significant predictor ($b = 0.045$, $t(73) = 2.47$, $p = .019$), but that there was a trending relationship for recognition memory with child-to-adult agreement ($p = .052$). Lastly, in our combined group model with adult-like segmentation as the outcome variable, we found that both age: older kids vs. adults ($b = -0.35$, $t(73) = 8.24$, $p < .001$) and age: younger kids vs. adults ($b = -0.44$, $t(73) = 6.17$, $p < .001$) were significant predictors of adult-like segmentation. Recognition memory was significantly related to adult-like segmentation as well ($b = -0.02$, $t(73) = 2.44$, $p = .017$).

We again conducted a post-hoc examination of the marginal means to break down the effects by all three age groups, and replicated our statistical findings for adults

compared to older and younger children, with the marginal means further suggesting that adults have more adult-like segmentation than older children ($b = 0.35$, $t(73) = 8.24$, $p < .001$) and younger children ($b = 0.44$, $t(73) = 6.17$, $p < .001$). There was no significant difference between older and younger children ($p = .072$) in adult-like segmentation. However, this is likely due to the variance accounted for by other independent variables in the model as well as the categorization of children into two distinct groups. When examining a simple correlation between continuous age and adult-like segmentation (i.e. child-to-adult agreement) in only children, we find that there is a significant correlation between age and adult-like segmentation in children ($r = .36$, $p = .022$; Figure 3B), suggesting that children's segmentation grows more adult-like with age across early childhood.

Verbal IQ

Because verbal IQ was a significant predictor of recognition memory for adults, as well as in our full group model, and was trending toward significance in several other models, we decided to explore this variable further. Interestingly, we found that verbal IQ was negatively correlated with age in our full sample including both adults and children ($r = -0.49$, $p < .001$). The KBIT-II is age-invariant because the scores are standardized, so this result indicates that children are outperforming adults on our measure of verbal IQ, once it has been standardized to account for age. To further investigate what might be driving this relationship, we asked whether parental education (of the children's parents, or of the parents of our adult participants, most of whom were undergraduate college students) was related to verbal IQ. Indeed, we found that in our full sample, performance on the KBIT-II Verbal was positively correlated with parental education ($r = .32$, $p =$

.005). A one-way ANOVA revealed a difference in parental education by age group (children vs. adults; $F_{(1,76)} = 27.36, p < .001$), with a post-hoc examination of the marginal means revealing that children's parents had greater levels of education than adults' parents in our sample ($b = 1.6, t = 5.23, p < .001$). Verbal IQ tasks have been shown to be skewed by socioeconomic variables such as level of parental education, but it is an unusual element of our study that age group would be confounded by this difference in verbal IQ performance. This phenomenon is likely because the undergraduate population at Temple University, from which our adult sample was primarily collected, is highly diverse, with students from a wide range of socioeconomic backgrounds, whereas many of our child participants were recruited from the suburbs of Philadelphia, where families are generally wealthy and highly educated.

Representational Similarity Analysis

Adult-Generated Boundary Results

For our first RSA, we used the adult-generated event boundaries to define the time windows we pulled from the data to be used in our analysis (described in Methods, Data Analysis, Imaging Data). To reiterate what was stated in the Methods, we chose the top ten most commonly identified boundaries and pulled time windows from either side of those (boundary comparisons), as well as from either side of pseudo-boundaries placed halfway between each boundary (within comparisons; Figure 6A). We then calculated dissimilarity values for every pair of time windows across the time series (40 total windows, one on each side of 10 boundaries and 10 pseudo-boundaries).

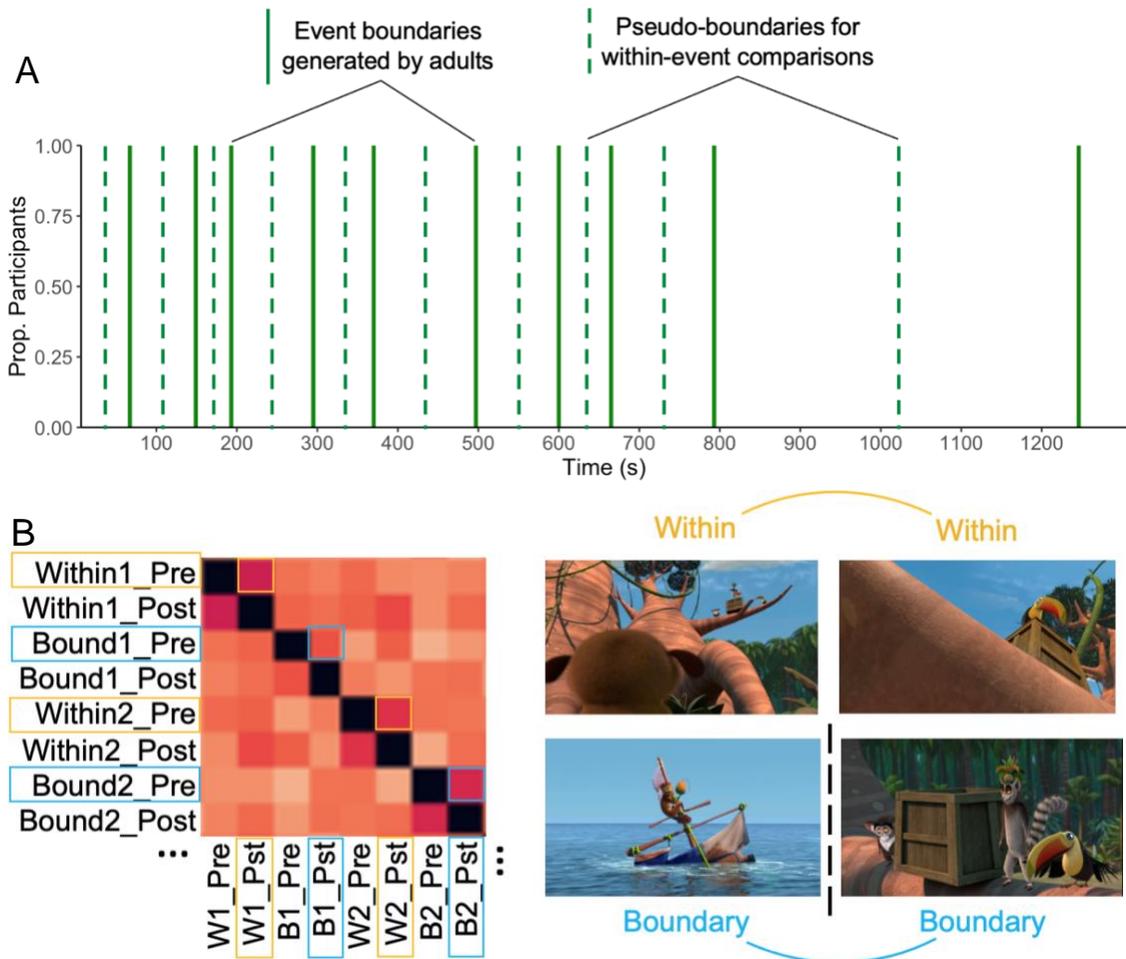


Figure 6. Across-boundary and within-event comparisons used in the RSA. A. Boundaries were generated by adults and children (adult boundaries depicted here), and pseudo-boundaries were placed halfway between each boundary. Across-boundary comparisons were based on time windows on either side of boundaries; within-event comparisons were based on windows on either side of pseudo-boundaries. B. For our RSA, pattern dissimilarity values were generated for every pair of time windows in the timeseries, but we used only the comparisons for time windows pulled from across one particular boundary or from within one given event.

The results of these comparisons can be visualized as a symmetrical matrix, an example portion of which is shown in Figure 6B. Our dissimilarity values of interest were the pairwise comparisons of the time windows on either side of a given boundary or pseudo-boundary (Figure 6B). These 20 dissimilarity values (one for each comparison; 10 boundary comparisons and 10 within comparisons) were extracted for each ROI, with

each value being labeled by whether it was an across-boundary or within-event comparison. We used these values to construct our models of interest.

For each of our six ROIs, we ran *t*-tests to compare the dissimilarity values for within-event comparisons vs. across-boundary comparisons. Our prediction was that, for the adult boundaries, there would be no difference between these two types of comparisons, because the children's neural data would not reflect the adult boundaries. For PCC, LAG, RAG, LHC, and RHC, there was indeed no significant difference between across-boundary and within-event dissimilarity values when using the adult-generated event boundaries (all *p*'s > .45). For EVC, there was a significant difference for boundary vs. within dissimilarity ($t(718) = 3.18, p = .002$; Figure 7), with the direction of the relationship being that dissimilarity for within-event comparisons was higher than for across-boundary comparisons, suggesting that pattern similarity was greater when viewing portions of events on either side of a boundary than from within the same event.

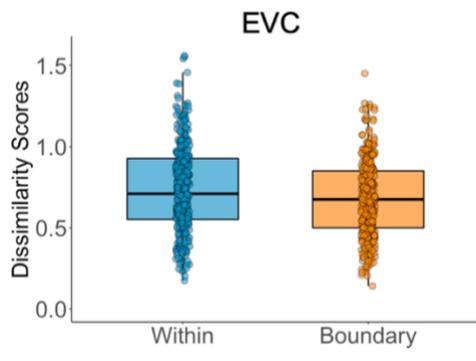


Figure 7. Pattern dissimilarity in EVC for adult-delineated boundaries. In EVC, pattern similarity was greater for across-boundary than within-event comparisons when the boundaries generated by adults were used to create the comparisons.

One possibility is that some of the adult boundaries were in locations where there was a change in action or characters present in the scene, but no scene change, leading to little change in perceptual features across a boundary. Further, a few adult boundaries were roughly centered around child boundaries, leading to the possibility that some of the

windows that were within an event for adults might have been near what children characterized as a boundary, leading to a possible explanation for the direction of this effect. In general, a lack of difference between across-boundary vs. within-event pattern dissimilarity in most ROIs for adult boundaries suggests that the children viewing the film in the scanner might not have perceived event boundaries in the locations at which our adult participants placed them.

Child-Generated Boundary Results

We conducted precisely the same analysis as we did with the 10 adult-generated boundaries with the 10 child-generated boundaries. Some of these child-generated boundaries were within a few seconds of adults' boundaries (Figure 2B, arrows), although only one was in exactly the same location. Others differed significantly in location from the adult-generated boundaries. Just as we did for the adult boundaries, we generated dissimilarity scores for every pairwise comparison of time windows across the timeseries and pulled the dissimilarity values for the relevant across-boundary and within-event comparisons from the resulting matrix (Figure 6).

We again ran t-tests comparing within-event to across-boundary dissimilarity scores for each ROI. We found a significant difference between within-event vs. across-boundary dissimilarity values for child-generated boundaries in PCC ($t(718) = 4.06, p < .001$), LAG ($t(718) = 4.53, p < .001$), RAG ($t(718) = 3.68, p < .001$), LHC ($t(718) = 4.06, p < .001$) and RHC ($t(718) = 2.13, p = .03$). For all of these regions, the mean for across-boundary dissimilarity was higher than for within-event dissimilarity (Figure 8), suggesting that neural activation patterns were more similar during time windows drawn from within a given event than during those drawn from either side of an event boundary.

There was a trending relationship for the difference between boundary and within comparisons for EVC ($p = .07$).

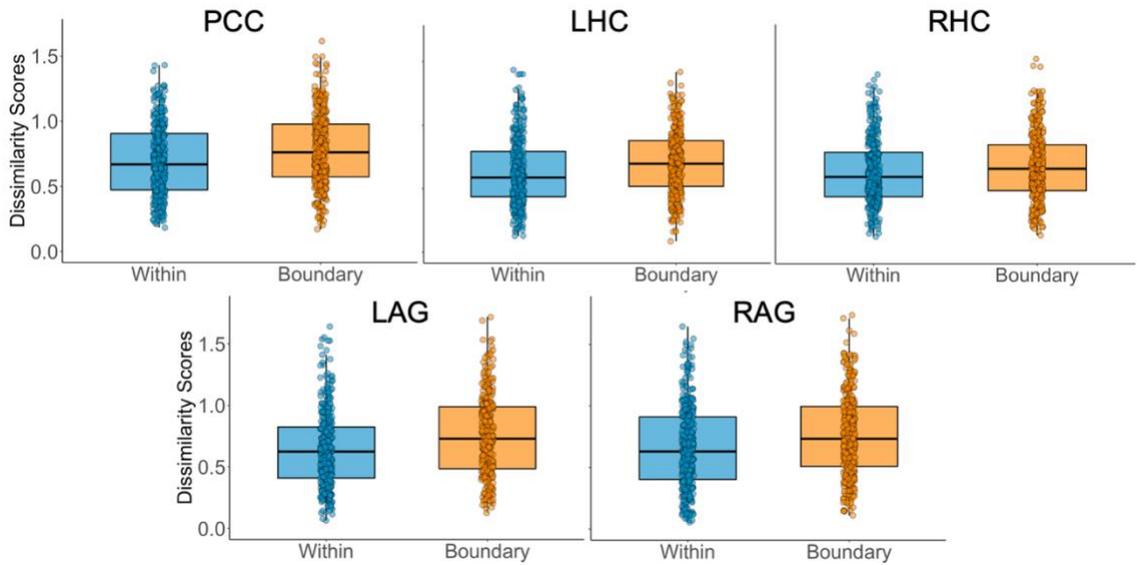


Figure 8. Pattern dissimilarity in PCC, LHC, LAG, and RAG for child-delineated boundaries. In these four regions, pattern similarity was greater for within-event than across-boundary comparisons when the boundaries generated by children were used to create the comparisons.

The significant results in these regions when using child-delineated boundaries suggests that event boundaries chosen behaviorally by children can be mapped onto a separate group of children’s neural data and will correspond to boundaries present in that data. Our hypothesis that this would be the case for boundaries behaviorally delineated by children but not adults is supported by our findings, illustrating that children likely track events similarly to each other, but might do so differently from adults.

Hidden Markov Model

As a way of verifying our results from a data driven perspective, we ran a Hidden Markov Model on our data in all ROIs (see Methods, Data Analysis, Imaging Data) to

determine whether the model would delineate boundaries in similar locations to where our child and adult participants placed boundaries. First, we denoised our preprocessed data by regressing out motion parameters, and then we z-scored the resulting output, which represented the entire timeseries, rather than specific time windows pulled from the timeseries, as in our RSA. This data was then used for all steps of the Hidden Markov Model. We defined the same six ROIs for this model as we did for our RSA. The model first calculated the correlation between each activation pattern within a given ROI across all 1605 TRs. Then, we asked the model to find 10 boundaries in the data for each ROI. In basic terms, the model steps through the data and tries to find the points at which stable activation patterns shift to new stable patterns; these are the locations at which the HMM places a boundary. The magnitude of correlation between patterns of activation at specific TRs varied with each ROI, as well as the locations of the boundaries and, consequently, the length of each event (Figure 9).

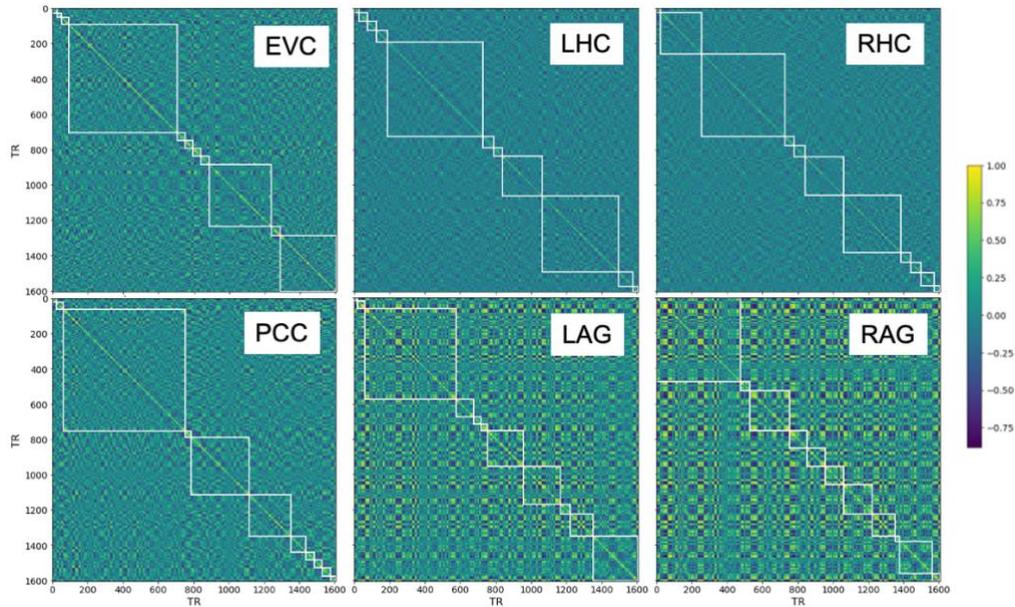


Figure 9. The 10 boundaries generated by the HMM for each ROI. When given the instructions to find 10 boundaries in the data, the HMM generated boundaries at different locations for each ROI. The matrix depicts correlations between activation patterns across the time series (shown here as 1605 TRs).

Next, we asked the model to determine the optimal number of boundaries for a given region. Prior work in adults has demonstrated that there is a hierarchical structure in the length of events determined by a given region, with lower-level regions such as EVC tending to be optimized for shorter events and, thus, more frequent event boundaries, and higher-level areas like PCC and AG being optimized for events of greater length, and consequently fewer total event boundaries. We found a similar pattern in our fMRI data in children (Figure 10), with the model detecting the most event boundaries/shortest events in EVC, fewer boundaries/longer events in PCC and AG, and still fewer boundaries/longer events in HC. These findings extend prior work in adults by showing that children’s brains track events hierarchically across different brain regions by detecting boundaries at more fine- or coarse-grained scales depending on the region.

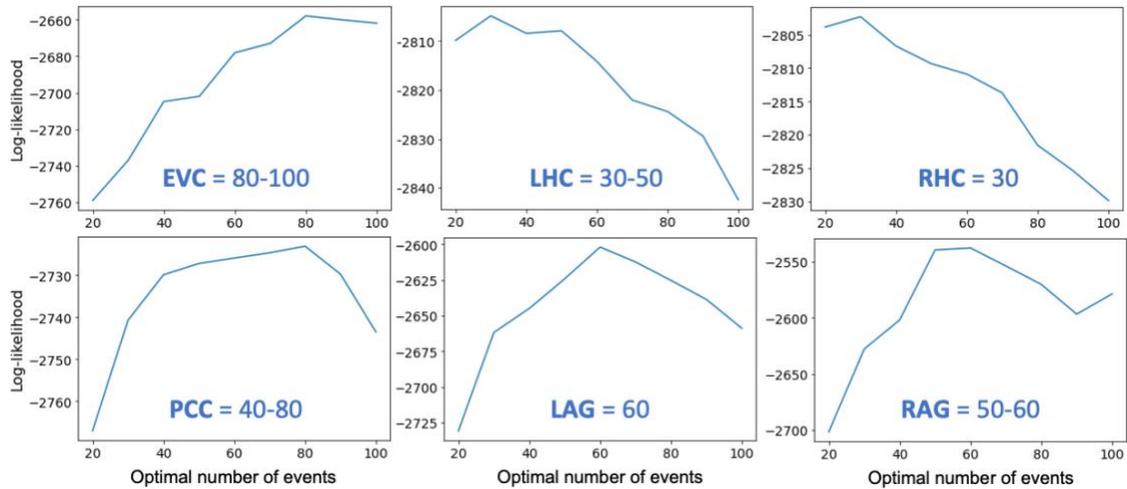


Figure 10. The optimal number of boundaries for each ROI. The HMM determined the optimal number of events per ROI, with the results suggesting a hierarchical structure for the ROIs' optimal number of events. Lower-level regions are optimized for more events on shorter timescales, and higher-level regions are optimized for fewer events on longer timescales.

Finally, we can ask whether the boundaries determined by the HMM are comparable to those chosen by behavioral participants. We chose the top 20 boundaries from both adults and children and converted them from seconds into TRs. We selected 20 boundaries to allow the model to fit to more data than the 10 boundaries used in our RSA, for which we had to keep the number of boundaries low to accommodate 32s on either side of each boundary. We then had the HMM define the locations of 20 boundaries in each ROI to match the number of boundaries defined by our human participants. We ran the traditional HMM, as well as a split-merge HMM (HMM-SM), which allows for more flexibility in the model to account for events differing greatly in length from one another across the timeseries (for details, see https://naturalistic-data.org/content/Event_Segmentation.html). We then conducted a test with 1,000 permutations that calculated the match between the boundaries detected by the HMM in each region and the boundaries delineated behaviorally by adults and by children. For this test, if there were two

boundaries in the same location between the HMM-determined boundaries and human-determined boundaries, or within 3 TRs (2.4 seconds) in either direction, this is considered a match. Matches were determined bidirectionally—that is, with the HMM boundaries as the ground truth and human boundaries as the proposed match, and vice versa.

With these criteria, we found that the HMM-delineated boundaries (both traditional and split-merge) were not significantly matched to the boundaries delineated by adults in any region (all p 's > .13). However, there was a significant relationship between the child delineated boundaries and the PCC HMM boundaries ($p = .041$), the EVC HMM-SM boundaries ($p = .042$), and the RAG HMM ($p = .027$) and HMM_SM ($p = .025$) boundaries. There were no other regions with significant matches to the child-delineated boundaries (all p 's > .11). The results of the matching analysis suggest that children's neural activity represents event boundaries at similar locations to where children place boundaries behaviorally in some regions. Additionally, the children's neural data does not represent the locations where adults behaviorally place boundaries. This provides support from a data-driven perspective for our findings from the RSA that children's brains track events in ways that map onto behavioral delineations of event boundaries, specifically those determined by other children rather than by adults.

CHAPTER 4

DISCUSSION

Prior work has shown that adults can easily segment ongoing naturalistic information such as movie clips into events (reviewed in Shin & DuBrow, 2021 and Zacks, 2020), and that patterns in adults' neural data reflect the boundaries they delineate behaviorally (Baldassano et al., 2017; Geerligs et al., 2021; Lee et al., 2020). With our study, we have shown that children can also segment naturalistic stimuli into events, although with greater variability and less consistency than adults. We have also shown that children who segment more like adults may be better able to remember the information from the events they segmented, and that children's segmentation gets more adult-like with development across early childhood. Additionally, we have demonstrated that children's neural activation patterns reflect boundaries delineated behaviorally by both adults and children, but is better represented by children's behavioral boundaries, suggesting that children may segment events in different locations than adults in early childhood. Together, these findings suggest that children's event cognition and memory is robust even in early childhood, but undergoes significant refinement across this developmental window and up to early adulthood.

Prior work has shown that children can segment events similarly to adults (Cohen & Baldassano, 2021; Zheng et al., 2020), but that their segmentation timing differs from adults' (Cohen & Baldassano, 2021) and is more variable (Zheng et al., 2020). We replicated these findings and showed that children as young as 4 can segment naturalistic stimuli into events. However, we also showed that children's segmentation gets more adult-like across early childhood as their event cognition develops and their perception of

ongoing stimuli likely more closely mirrors that of adults'. Although they show some agreement with the adult norm, children's segmentation is less consistent than adults' and they have more variability in their choice of boundaries, which we showed both qualitatively and quantitatively. The magnitude of adults' agreement with one another was generally much higher than that of children, but children still showed substantial agreement with each other in the placement of event boundaries. The within-group segmentation agreement for children also reflects that they have strong event cognition and that their understanding of events is shared with one another, in ways that are both like adults and different from adults.

This differential between children's and adults' segmentation in turn maps onto the neural data of a separate group of children in the same age range. We have shown that boundaries delineated behaviorally by one group of children can be mapped onto a separate group of children's neuroimaging data. Although our RSA found one region in which adults' boundaries corresponded to the neural data, the children's boundaries fit well consistently in the case of the RSA and much better in terms of the HMM, suggesting that children may perceive event boundaries in different locations than adults do. Indeed, just over half of children's boundaries were similar to adults', while the other half were in disparate locations. Even in the places where children "agreed" with adults, the locations often differed by several seconds, which could be enough to be reflected in the neural activation windows chosen for our RSA, since the windows were placed only 3 seconds from a given boundary and had only a 5 second duration. It may be that children define the ends of some events similarly to adults, but they may in some cases determine that a boundary comes earlier or later in time than adults do, which could be reflected in

the timing of shifts in neural pattern activity. Overall, it seems that children have a good understanding of events and their boundaries, but that they may define boundaries differently than adults in some cases, and their ability to consistently track events across an extended period of time is still developing. The implications of this are that children group contexts, objects, and people together in their experiences in ways that might not reflect adults' groupings, which could in turn affect how children recall their experiences and form semantic knowledge and schemas. Prior research suggests that children can rapidly reorganize their semantic knowledge after naturalistic learning experiences (Unger & Fisher, 2019), suggesting that what children experience and how they perceive it influences the long term knowledge they form. The two are likely reciprocal, with children's perception of ongoing events being influenced by their existing semantic knowledge, and the development of new semantic knowledge in turn being influenced by how they experience the temporal unfolding of their daily life and group it into events.

In addition to children's perception of events, their memory for events is still developing across childhood as well. Although children's recognition memory was good even at the youngest ages in our sample, older children still outperformed younger children, who were yet outperformed by adults. This replicates prior work showing a similar stepwise improvement across these age groups on episodic memory tasks (Benear et al., 2021; Ngo, Benear, Popal, Olson, & Newcombe, 2021; Ngo et al., 2018). Adults, on the other hand, were nearly at ceiling on this task, suggesting that even though children can perform above chance on this recognition task, and in some cases perform quite well, they are not yet at adult-like levels of memory, especially at ages 4 and 5. For temporal order memory, older children performed well, but younger children did not

perform above chance level, suggesting that performance on temporal order tasks undergoes significant developmental change in this age window and stabilizes later than recognition memory, which is supported by prior research (Bettencourt, Everett, Chen, & Pathman, 2021; Price & Goodman, 1990; Blankenship, Calkins, & Bell, 2022). Older children and adults were able to perform above chance on our temporal order memory task, but even adults had some variability in performance. Recalling the temporal order of events has been shown even in adults to be a more challenging task than simple recognition, with dissociations becoming clear in aging populations (Fabiani & Friedman, 1996), which aligns with our findings on the differential performance for recognition compared with temporal order memory in our study, especially in children.

Prior work has shown that brain activation patterns can track naturalistic stimuli by representing within-event stimuli differently than across-event stimuli in adults (Chanales, Oza, Favila, & Kuhl, 2017) and in children (Benear, Horwath, et al., in press), and that children's brains represent event boundaries (Cohen & Baldassano, 2021). Here we extend these findings by showing that we can use boundaries delineated by children to define what is considered an event, and that other children's brains detect boundaries at these same locations, which differ from the locations chosen by adults. This is a crucial finding, because it suggests that even very young children are grouping their ongoing perceptual experience into events, but that they may do so in a qualitatively different way than adults do. One thing we noticed when examining children's event segmentation data was that children tended to place boundaries in locations where a highly emotional, visible, or entertaining action took place, such as a character flying across the screen or something blowing up, whether or not it signified the end of an event as defined by

adults. This could be related to prediction error, which has been shown to be tied to behavioral demarcations of event boundaries by adults (reviewed in Zacks, 2020). When prediction error is high due to a sudden or unexpected movement or sound in an ongoing stream of perceptual input, children may perceive this as a new event, even though the same characters may still be present in the same scene, for example. Violations of one's expectations in the real world are often correlated with event boundaries, so children may associate the two even when they're decoupled. Prior work has demonstrated that children's attention at event boundaries increases (Meyer, Baldwin, & Sage, 2011), indicating that even very young children respond to the salience of event boundaries by paying attention to changes in their environment, whether it be the arrival of new people, a change of location, or the introduction of a threat. Of course, what marks the end of one event and the beginning of a new one has some objective metrics, but is still a subjective judgment, so it may be that different markers signal new events to children than to adults. Children have encountered less of the world than adults have and may experience more frequent and impactful violations of their expectations because they have fewer schemas for how situations tend to unfold. Thus, they may perceive event boundaries in different locations than adults do under some conditions.

Our neural findings are relatively novel, with only one other study we are aware of evaluating children's event cognition and comparing it to their behavioral demarcations of event boundaries (Cohen & Baldassano, 2021). Similarly to this study's findings, we found that a Hidden Markov Model was able to find event boundaries in children's fMRI data in several regions implicated in event processing, and we have replicated those findings with our HMM and extended the findings with our

representational similarity analysis. Our RSA results show that children's brains represent time windows drawn from the same event more similarly than those drawn from either side of an event boundary, which has profound implications for our understanding of how children process the world around them. Although children's ability to recall events is still improving across early childhood, they are still processing events as events during perception, supported by evidence from both our RSA and HMM that there are robust boundaries in children's neural data while they engage with a naturalistic stimulus such as a television show. However, this event processing might be fundamentally different from adults', since adults' boundaries did not align as well with the children's neural data in either analysis.

Our RSA demonstrated that children's behavioral boundaries could be mapped onto children's neural data in the PCC and AG, regions that have been shown to track events in adults (Baldassano et al., 2017; Geerligs, Van Gerven, Campbell, & Güçlü, 2021; Lee, Aly, & Baldassano, 2020). These regions likely recur as relevant to event processing because they are part of parietal cortex, which research has shown is tied to recalling events in the correct temporal order and representing magnitudes, including of time and space (Bueti & Walsh, 2009; Godbout, Cloutier, Bouchard, Braun, & Gagnon, 2004). The RSA results also showed that HC differentially represented within-event and across-boundary details for children's behaviorally-delineated boundaries, indicating a potential role for this region in offloading events into long term memory for children, as previously shown in adults (reviewed in Zacks, 2020). Using the HMM, shifts in brain states signaling event boundaries in the neural data that aligned with the behavioral data were present only in PCC, RAG, and EVC, but not in HC. Our RSA findings suggest that

the hippocampus may represent stimuli that share overlapping features more similarly than stimuli that do not, but the HMM results suggest that it may not demonstrate the same state changes over time as in the perceptual and higher level regions in which the HMM detected boundaries. Both the RSA and HMM results determined that there were event boundaries in the PCC and RAG that aligned with the children's behavioral boundaries, suggesting that these two higher-level regions implicated in event processing in adults serve a similar role in children. The HMM showed a significant relationship between children's behavioral and neural boundaries in EVC, but the relationship was only significant for the RSA only when adults' boundaries were used, and in the opposite direction. Early visual cortex tracks perceptual features, which shift much more quickly than higher-level event changes, supported by the HMM's output for the optimal number of events in this region, which was much higher than in the other ROIs. It may be that, because of these frequent shifts, it's harder to map behavioral boundaries on longer timescales onto this region in a way that is consistent and meaningful. However, this region is clearly implicated in event processing in children, with shifts in brain states tracking event boundaries to some degree. Future research should collect fMRI data from both adults and children, as well as collecting both neural and behavioral data from all participants, to make comparisons across groups and within individuals in terms of how behavioral boundaries might differentially relate to neural data across development.

In conclusion, children have a rich, complex understanding of events that is reflected in their ability to demarcate boundaries and recall parts of events quite effectively even at very young ages. However, their event cognition grows more adult-like across this developmental window, suggesting that their understanding of events is

being refined. Children's event memory is also good even at age 4, with recognition memory being predicted by more adult-like segmentation, although temporal order memory in particular improves across the age range. Finally, signatures of event boundaries can be detected in children's neural data, are organized hierarchically by ROI as in adult data, and map well onto children's behaviorally-delineated boundaries. Altogether, our findings demonstrate that children's event cognition is robust from at least the age of 4 both behaviorally and neurally, but that children's tendency to demarcate event boundaries in the same manner as adults increases across early childhood, as their memory for events also improves.

REFERENCES CITED

- Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5.
- Ben-Yakov, A., & Henson, R. N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *Journal of Neuroscience*, *38*(47), 10057–10068.
- Beneat, S. L., Ngo, C. T., Olson, I. R., & Newcombe, N. S. (2021). Understanding relational binding in early childhood: Interacting effects of overlap and delay. *Journal of Experimental Child Psychology*, *208*, 105152.
- Bettencourt, K. M., Everett, L. H., Chen, Y., & Pathman, T. (2021). Examining the development of memory for temporal context and its underlying neural processes using event-related potentials. *Developmental Cognitive Neuroscience*, *48*, 100932.
- Blankenship, T. L., Calkins, S. D., & Bell, M. A. (2022). The Role of Executive Functions in Item Recognition and Temporal Order Memory Development. *Journal of Cognition and Development*, *23*(1), 135–147.
- Bloom, P. (2002). *How children learn the meanings of words*. MIT press.
- Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1525), 1831–1840.
- Carey, S. (1982). Semantic Development: The State of the Art. In E. Wanner & L. R. Gleitman (Eds.), *Language Acquisition: The State of the Art* (pp. 347–387). Cambridge: Cambridge University press.

- Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, 27(15), 2307-2317.e5.
- Cohen, S. S., & Baldassano, C. (in press). Developmental changes in story-evoked responses in the neocortex and hippocampus. *eLife*.
- Daugherty, A. M., Flinn, R., & Ofen, N. (2017). Hippocampal CA3-dentate gyrus volume uniquely linked to improvement in associative memory from childhood to adulthood. *NeuroImage*, 153(December 2016), 75–85.
- Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., ... Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, 16(1), 111–116.
- Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, 81(5), 1179–1189.
- Fabiani, M., & Friedman, D. (1996). Dissociations between memory for temporal order and recognition memory in aging. *Neuropsychologia*, 35(2), 129–141.
- Farrar, M. J., & Goodman, G. S. (1992). Developmental Changes in Event Memory. *Child Development*, 63(1), 173–187.
- Geerligs, L., Van Gerven, M., Campbell, K. L., & Güçlü, U. (pre-print, 2021). A nested cortical hierarchy of neural states underlies event segmentation in the human brain. *BioRxiv*, 2021.02.05.429165.
- Godbout, L., Cloutier, P., Bouchard, C., Braun, C. M. J., & Gagnon, S. (2004). Script generation following frontal and parietal lesions. *Journal of Clinical and*

- Experimental Neuropsychology*, 26(7), 857–873.
- Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., & Ghosh, S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*.
- Hudson, J. A., Fivush, R., & Kuebli, J. (1992). Scripts and episodes: The development of event memory. *Applied Cognitive Psychology*, 6(6), 483–505.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2), 782–790.
- Keresztes, A., Bender, A. R., Bodammer, N. C., Lindenberger, U., Shing, Y. L., & Werkle-Bergner, M. (2017). Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. *Proceedings of the National Academy of Sciences*, 114(34), 9212–9217.
- Keven, N. (2016). Events, narratives and memory. *Synthese*, 193(8), 2497–2517.
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2), 72–79.
- Lee, C. S., Aly, M., & Baldassano, C. (2020). Anticipation of temporally structured events in the brain. *BioRxiv*, 1–18. <https://doi.org/10.1101/2020.10.14.338145>
- Lee, C. S., Aly, M., & Baldassano, C. (2021). Anticipation of temporally structured events in the brain. *ELife*, 10, 1–15.
- Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, 36(3), 273–293.
- Meyer, M., Baldwin, D. A., & Sage, K. (2011). Assessing Young Children’s Hierarchical

- Action Segmentation. *How to Use Mechanical Turk for Cognitive Science Research*, 66–67.
- Moore, E. M., Glass, L., Infante, M. A., Coles, C. D., Kable, J. A., Jones, K. L., ... CIFASD, the. (2021). Cross-Sectional Analysis of Spatial Working Memory Development in Children with Histories of Heavy Prenatal Alcohol Exposure. *Alcoholism: Clinical and Experimental Research*, 45(1), 215–223.
- Newcombe, N. S., Lloyd, M. E., & Ratliff, K. R. (2007). *Development Of Episodic And Autobiographical Memory: A Cognitive Neuroscience Perspective. Advances in Child Development and Behavior* (Vol. 35). Elsevier Inc.
- Ngo, C. T., Benear, S. L., Popal, H., Olson, I. R., & Newcombe, N. S. (2021). Contingency of semantic generalization on episodic specificity varies across development. *Current Biology*, 31, 1–8.
- Ngo, C. T., Lin, Y., Newcombe, N. S., & Olson, I. R. (2019). Building Up and Wearing Down Episodic Memory: Mnemonic Discrimination and Relational Binding. *Journal of Experimental Psychology: General*.
- Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental Science*, 21(2), 1–11.
- Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2019). Gain-Loss Framing Enhances Mnemonic Discrimination in Preschoolers. *Child Development*, 90(5), 1569–1578.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ...

- Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.
- Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M.-L., ... Kiselev, V. G. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, *42*(1), 87–97.
- Price, D. W. W., & Goodman, G. S. (1990). Visiting the Wizard: Children's Memory for a Recurring Event. *Child Development*, *61*(3), 664–680.
- Ren, J., Wharton-Shukster, E., Bauer, A., Duncan, K., & Finn, A. S. (2021). Events structure information accessibility less in children than adults. *Cognition*, *217*, 104878.
- Riggins, T., Geng, F., Botdorf, M., Canada, K., Cox, L., & Hancock, G. R. (2018). Protracted hippocampal development is associated with age-related improvements in memory during early childhood. *NeuroImage*, *174*(January), 127–137.
- Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., ... Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241–255.
- Sasmita, K., & Swallow, K. M. (2022). Measuring event segmentation: An investigation into the stability of event boundary agreement across groups. *Behavior Research Methods*.
- Shin, Y. S., & DuBrow, S. (2021). Structuring Memory Through Inference-Based Event Segmentation. *Topics in Cognitive Science*, *13*(1), 106–127.
- Unger, L., & Fisher, A. V. (2019). Rapid, experience-related changes in the organization of children's semantic knowledge. *Journal of Experimental Child Psychology*, *179*,

Yates, T. S., Skalaban, L. J., Ellis, C. T., Bracher, A. J., & Turk-Browne, N. B. (2021).

Neural event segmentation of continuous experience in human infants, 1–27.

Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, *71*,

165–191.

Zheng, Y., Zacks, J. M., & Markson, L. (2020). The development of event perception and

memory. *Cognitive Development*, *54*, 100848.