Received: 14 October 2021 | Revised: 26 January 2023 | Accepted: 12 April 2023

DOI: 10.1111/desc.13410

Developmental Science WILEY

RESEARCH ARTICLE

Staying and Returning dynamics of young children's attention

Jaeah Kim[1] | **Shashank Singh**[2] | **Catarina Vales**[1] | **Emily Keebler**[1] | **Anna V. Fisher**[1] | **Erik D. Thiessen**[1]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany

Correspondence
Jaeah Kim, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.
Email: jaeahk@andrew.cmu.edu

Funding information
National Science Foundation, Grant/Award Number: BCS-1451706; Institute of Education Sciences, Grant/Award Number: R305B150008

Abstract

In this paper, we decompose selective sustained attending behavior into components of continuous attention maintenance and attentional transitions and study how each of these components develops in young children. Our results in two experiments suggest that changes in children's ability to return attention to a target locus after distraction ("Returning") play a crucial role in the development of selective sustained attention between the ages of 3.5–6 years, perhaps to a greater extent than changes in the ability to continuously maintain attention on the target ("Staying"). We further distinguish Returning from the behavior of transitioning attention away from task (i.e., becoming distracted) and investigate the relative contributions of bottom-up and top-down factors on these different types of attentional transitions. Overall, these results (a) suggest the importance of understanding the cognitive process of transitioning attention for understanding selective sustained attention and its development, (b) provide an empirical paradigm within which to study this process, and (c) begin to characterize basic features of this process, namely its development and its relative dependence on top-down and bottom-up influences on attention.

KEYWORDS
attention development, attention measurement, attention over time, attentional transitions, eye-tracking, selective sustained attention, TrackIt

Research Highlights

- Young children exhibited an endogenously ability, Returning, to preferentially transition attention to task-relevant information over task-irrelevant information.
- Selective sustained attention and its development were decomposed into Returning and Staying, or task-selective attention maintenance, using novel eye-tracking-based measures.
- Returning improved between the ages of 3.5–6 years, to a greater extent than Staying.
- Improvements in Returning supported improvements in selective sustained attention between these ages.

1 | INTRODUCTION

In an ever more information-dense society, the ability to sustain one's attention on relevant information to complete tasks is crucial. Many everyday tasks require extended durations of effort to successfully perform or make progress on; key examples include learning, driving, writing, and playing sports (Bennett et al., 1976; Harris et al., 2017; Kent et al., 2014; Monk et al., 2004; Rosenshine & Berliner, 1978). Over the course of maintaining effort on a task, our attention may regularly visit other task-irrelevant loci, whether due to external (Parmentier & Andrés, 2010) or internal (Smallwood & Schooler, 2015) distractions or due to fatigue or drowsiness (Thomson et al., 2015). Furthermore, few real-life settings provide external prompts to guide attention back to the task when distracted. Thus, successfully sustaining attention on a task may require not only the ability to continuously maintain attention on the task for prolonged periods of time, but also the ability to return one's attention to the task, in a self-initiated manner, after distractions. While both of these component abilities develop between infancy and adulthood, their distinct developmental trajectories are unknown, and it is unclear which, if not both, of these behaviors underlie pronounced developments in selective sustained attention that are known to occur during early childhood (Diamond, 2006; Ruff & Rothbart, 2001). Moreover, it is not known what common or distinct cognitive mechanisms underlie these two abilities.

In general, attending over time consists of continuous periods of attention maintenance on a single locus, which we will refer to as *attention maintenance*, delimited by *attentional transitions* from a *source* locus to a *destination* locus. This allows for at least two mechanisms by which attention could support successfully completing a task, which requires selectively attending more to task-relevant information than to task-irrelevant information. First, selectivity could influence attention maintenance; this would result in attending to task-relevant information *for longer durations* than to task-irrelevant information, a behavior we refer to as *Staying*. Second, selectivity could influence attentional transitions; this would result in transitioning attention to task-relevant information *more frequently* than to task-irrelevant information, a behavior we refer to as *Returning*. These selective mechanisms are not necessarily mutually exclusive. While attention maintenance and transition behaviors have been distinguished before (Cohen, 1969, 1972; Posner & Cohen, 1984), our notions of Staying and Returning specifically reflect the distinct supportive roles these behaviors may play in attending to task-relevant information.

Attention can be guided towards task-relevant information by exogenous factors (e.g., increased relative salience of task-relevant information) or endogenous factors (e.g., preference for information relevant to one's internal goals). In settings lacking exogenous support for attending to the task, and therefore necessitating endogenous support for attending to the task, Staying and Returning represent two possible routes of endogenous control in selective sustained attention. In particular, Staying might rely on endogenous control when determining whether to terminate or maintain attentional engagement with the current locus based on its task relevance. Meanwhile, Returning might rely on endogenous control when searching for and shifting attention to a task-relevant destination locus (Brown & Denney, 2007; Lamy & Egeth, 2002). We note that, while attentional loci can be distinguished at many levels of granularity (e.g., features, spatial locations, objects, etc.), and the notions of Staying and Returning are equally applicable across these granularities, this paper focuses on object-based attention, in which a "locus" refers to a single object on which an individual's attention is placed (Baylis & Driver, 1992; Egly et al., 1994).

In this work, we study the role of endogenous control in selective sustained attending through novel eye-tracking measures of selective attention maintenance (Staying) and selective transitioning (Returning), as well as development of these behaviors between the ages of 3.5 and 6 years and their roles in the development of selective sustained attention during these years. Before describing specific goals of this research, we provide relevant background on the development of endogenous control of attention.

1.1 | Development of endogenous control of sustained attention

Early in development, attention is believed to be driven primarily by exogenous factors (Colombo & Cheatham, 2006). In a classic series of studies, Cohen (1969, 1972) decomposed looking behaviors of 4-month old infants presented with a single stimulus into "attention-holding" and "attention-getting" component behaviors, which can be thought of as analogs of our Staying and Returning behaviors, respectively, in the simpler setting where only one stimulus is presented. Cohen then investigated how several exogenous factors influence attention-holding and attention-getting. Attention-getting, measured by latency of the infant's first look to the stimulus after onset, was found to be sensitive to the size and salience of the stimulus rather than its novelty or complexity, and was therefore believed to reflect an orienting process (Cohen et al., 1975). Meanwhile, attention-holding, measured by duration of an infants' first look to the stimulus, was found to be sensitive to the novelty and complexity of the stimulus and was, thus, hypothesized to reflect speed of information processing; for example, attention-holding exhibits habituation effects, decreasing in response to repeated exposures to the same stimulus (Cohen et al., 1975).

Studies of looking times in infants and toddlers suggest that endogenous control over sustained attention begins to emerge late in the first year of life, reflected in increases in both overall look durations and the latency of looking to distractors (Colombo & Cheatham, 2006). While these looking time effects plateau through the second and third years of life, in other settings, such as free play with novel toys, duration of attention continues to increase throughout the preschool years (Moyer & Gilmer, 1955; Ruff & Rothbart, 2001). The ability to endogenously focus attention towards endogenously interesting loci, such as human faces, in the presence of salient but qualitatively different distractors is clearly present in the the first year of life (Holmboe et al., 2008; Kwon et al., 2016). However, the ability to endogenously focus attention towards a Target, in the presence of distractors, that is distinguished *only by its relevance to an explicit task or goal*, which is the focus of the present study, emerges later, at approximately 3.5 years, and continues to develop throughout early childhood (Ruff & Capozzoli, 2003).

This ability supports selective sustained attention (SSA)[1], the ability to enhance the processing of task-relevant information at the exclusion of task-irrelevant information over a period of time (Fisher, 2019). SSA is in turn crucial to performing tasks of executive control, at which children improve significantly between the ages of 3.5 and 6 years (Diamond, 2006; Fisher, 2019). Improvements in selectively sustaining attention to task-relevant information have been measured in multiple studies; for example, Deng and Sloutsky (2016) and Plebanek and Sloutsky (2017) have shown that, while children's abilities to detect changes in task-relevant information improve within this age range, their ability to detect changes in *task-irrelevant* information may degrade at the same time. These results accentuate increased *selectivity* of processing and suggest a developmental trend along the exploration–exploitation trade-off (the competition between maximizing short-term performance and longer-term learning) in the guidance of attention (Blanco & Sloutsky, 2020; Dubois et al., 2020; Gopnik, 2020; Laureiro-Martínez et al., 2010). Within the same SSA task paradigm used in the present study, Fisher et al. (2013) showed that children's performance in tracking a target object amongst equally salient distractors improves between the ages of 3 and 6 years, approaching performance in an easier exogenously supported condition in which the target is more salient than distractors. This suggests that developmental improvements in task performance during these years reflect an increasing contribution of endogenous factors in the control of SSA. In a task where children classified animals into categories based on their habitat, Wetzel et al. (2019) found that children responded with a delay on trials preceded by novel task-irrelevant distractor sounds, and that this delay decreased with age, most prominently between the ages of 4 and 6 years, similarly suggesting maturation of endogenous attentional control during these years.

These studies show age-related improvements in SSA between the ages of 3 and 6, suggesting this to be a period of rapid change in the endogenous ability to maintain attention to the task at hand in the presence of distractors. However, it is still unclear whether this occurs via Staying and/or Returning behaviors, influencing either the duration or frequency of attending to task-relevant information relative to task-irrelevant information. To address this gap, the present paper examines which of these component behaviors mediates developmental improvements in young children's endogenous control of their SSA.

[1] We avoid the more common term "sustained attention" because it has been used with distinct meanings in different literatures. On one hand, "sustained attention" has been used to refer to continuous maintenance of attention on a single locus (without any attentional transitions), independent of any particular task or goal. This definition is especially common in the context of infants (Colombo & Cheatham, 2006). On the other hand, in the context of participants performing a goal-oriented task (such as TrackIt), "sustained attention" has been used to refer to sustained *selectivity* of attention (i.e., enhanced processing of task-relevant information over task-irrelevant information, maintained over an extended duration; Egeland & Kovalik-Gran, 2010), which can be supported by selective mechanisms such as Staying and Returning. We refer to this latter notion as "selective sustained attention," as suggested by Fisher and Kloos (2016).

1.2 | Current studies

The studies reported here aim to address the following questions. Can we identify and measure task-selective attention maintenance (Staying) or transitioning (Returning) behaviors within SSA? How are Staying and Returning behaviors influenced by exogenous and endogenous factors? In particular, which, if not both, of these behaviors exhibits endogenous selectivity for task-relevant information? How do Staying and Returning change with development? In particular, which, if not both, of these behaviors changes to support increases in attention selectivity with development between the ages of 3 and 6 years?

This paper's first contribution is to develop measures of Staying and Returning, described in Section 2.1.2, based on gaze data collected while children performed TrackIt, a visual object tracking task in which participants track a single moving Target object in the presence of multiple moving Distractor objects (Fisher et al., 2013). TrackIt performance has been shown to have good psychometric properties as a measure for SSA and has been linked to numeracy skills, prospective memory, proactive control, language development, and learning in classroom-like settings (Brueggemann & Gable, 2018; Doebel et al., 2017, 2018; Erickson et al., 2015; Fisher et al., 2013; Kim et al., 2017; Mahy et al., 2018; Smolak et al., 2020). Moreover, TrackIt is designed so that participants' gaze is closely coupled to their attentional state over time. The high spatiotemporal resolution of gaze data, thus, allows us to precisely identify periods of continuous maintenance of attention on both Target and Distractor objects, as well as transitions of attention between them, using a method recently developed by Kim et al. (2020) for inferring participants' internal attentional states from their gaze data. Importantly, while children perform TrackIt, no external cues or prompts are provided to direct attention. Additionally, within each TrackIt trial, Target and Distractor objects are sampled randomly (without replacement) from a common pool of objects, such that the Target and Distractors do not systematically differ in terms of relative salience, novelty, or complexity, and prior work also suggests that children's familiarity with the Target shape does not influence their TrackIt performance (Keebler et al., 2019). Thus, any systematic preference we observe for attending to the Target, whether through Staying or Returning, should reflect the influence of endogenous factors, based on children's internal representations of the task goal.

We then utilize these new measures to investigate the cognitive and developmental roles of Staying and Returning behaviors. In Experiment 1, we first use our new measures to investigate whether children exhibit Staying and Returning behaviors (in the forms of preferential attention maintenance and transitioning, respectively, towards the Target object). Next, we investigate developments in Staying and Returning over age, as well as their role in TrackIt task performance. Finally, we perform a post hoc investigation of a possible exogenous influence on attentional transitions between objects. To strengthen external validity of our results and provide confirmatory analyses for our post hoc analyses, Experiment 2 replicates these analyses on a second independently collected dataset. Additionally, since our results rely on novel automated methods for analyzing eye-tracking data, Appendix A.2 experimentally validates these novel methods by

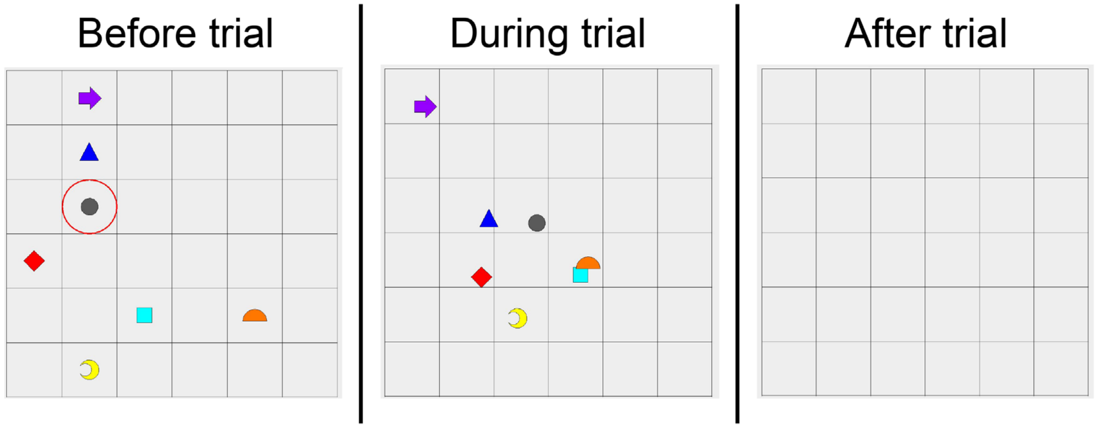

FIGURE 1 An example trial of the standard TrackIt task. The Target object here is the gray circle, as indicated before the trial by a red circle. During a trial, all objects move in unpredictable piecewise-linear paths and disappear after a set duration. After a trial, participants are asked to indicate the grid cell where the Target was before disappearing. A video of an example TrackIt trial provided by Kim et al. (2020) can be found at https://github.com/CMU-CDL/TrackIt/blob/main/endogenous_TrackIt_example.mp4?raw=true.

replicating our results using eye-tracking data hand-coded by trained human coders.

2 | EXPERIMENT 1

2.1 | Methods

2.1.1 | TrackIt task

TrackIt is a visual object-tracking task introduced by Fisher et al. (2013) to measure SSA in young children. TrackIt allows developmentally sensitive assessment over a range of ages, with children as young as 3 years old consistently completing the task and providing usable data (Fisher et al., 2013; Kim et al., 2017). In the TrackIt task (illustrated in Figure 1), participants are instructed to track, using only their eyes, a single *Target* object moving about on a grid, among other moving *Distractor* objects. At the end of each trial, all objects vanish from the grid, and participants are asked to identify the grid cell the Target occupied immediately before vanishing. The accuracy of this final response, referred to as *Location Accuracy*, is used as the main behavioral measure of task performance.

After providing a Location response, participants are presented with four distinct objects, taken from the objects displayed during the trial, one of which was the Target object in that trial. The accuracy with which participants identify the Target object, called *Memory Accuracy*, has been used as a secondary performance measure to help distinguish Location errors due to forgetting, or otherwise being unaware of the Target object, from Location errors due to other attention failures.

In this study, we used data from children performing TrackIt together with gaze data collected while children performed TrackIt. Because TrackIt requires continuous overt attention to the Target, eye-tracking provides information about a participant's visual attention with high temporal resolution. In particular, leveraging an algorithm introduced by Kim et al. (2020) to infer the object of a participant's attention from their eye-tracking data, we are able to identify, at high frequency and with high confidence, which object a participant is attending to at each point in time. These methodological features make TrackIt uniquely suited to investigating how attention moves over time, in contrast to other widely used SSA tasks, such as the continuous performance test (CPT), which collect data and require participant attention at most once every few seconds (Fisher & Kloos, 2016; Riccio et al., 2002; Rosvold et al., 1956). Furthermore, the fact that TrackIt explicitly provides task-irrelevant objects, alongside the Target, to which the participant can attend is important because it allows us to isolate the effect of task-relevance on the participant's attending behavior. Finally, prior research has shown that TrackIt is more reliable for use with young children than tasks such as the CPT and the multiple object tracking (MOT) task[2], in which young children often fail to complete the task or fail to meet minimum performance criteria for their data to be included (Akshoomoff, 2002; Brockhoff et al., 2016; Fisher & Kloos, 2016).

2.1.2 | Data

We analyze a dataset of TrackIt and eye-tracking data originally collected by Kim et al. (2020), publicly available on the Open Science Framework (DOI 10.17605/OSF.IO/U8JBS) at https://osf.io/u8jbs/. Python code for reproducing our analyses is available on GitHub at https://github.com/Jaeah/staying-and-returning. Here, we briefly review the data collection process of Kim et al. (2020).

2.1.3 | Participants

Kim et al. (2020) recruited 50 typically developing children, aged 3.5–6 years ($M = 4.60$ years, $SD = 0.67$ years), 23 male and 27 female,

[2] TrackIt is closely related to MOT, a task in which participants track *multiple* Target objects simultaneously among *identical* Distractor objects; Appendix A.5 discusses some relevant connections between TrackIt and MOT.

from a laboratory school affiliated with an urban private university in the northeastern United States. Race and ethnicity information for the sample reported by parents or guardians was as follows: 62% White, 4% Black or African American, 10% East Asian or Asian American, 4% Native American or Alaskan Native, 2% South Asian or Indian American, 2% Jewish, and 14% unreported.

2.1.4 | Procedure

Each child performed 11 TrackIt trials, as described above under "TrackIt Task," including one initial practice trial during which the experimenter explained the task. Practice trials were not analyzed, giving 10 usable trials per participant. The experimental protocol was approved by the University Institutional Review Board, and informed consent was obtained from a parent or guardian of each participant.

2.1.5 | TrackIt parameters

We used version 3.1.0 of TrackIt, available at https://sites.google.com/andrew.cmu.edu/trackit/home/task-versions, which has several parameters that can be adjusted to modulate task difficulty. Kim et al. (2020) used parameter values identified as age-appropriate for 4- to 6-year-old participants based on prior research by Kim et al. (2017). Specifically, they set the number of Distractors to 6, grid size to 6×6, and object speed to 500 pixels/s ($\approx 10°$/s). For each trial, Target and Distractor objects were sampled without replacement from nine distinct shapes, and their colors were sampled without replacement from nine distinct colors; that is, out of 81 possible shape-color combinations, Target and Distractor objects were selected randomly under the constraint that all objects within a trial had distinct shapes and colors. The minimum trial length was set to 10 s, although exact trial length varied between 10.1 and 27.5 s ($M = 14.4$ s, $SD = 3.0$ s) to reduce predictability of the timing of trial end. All other parameters, such as display frequency (30 Hz) and motion interpolation ("Linear"), were set to TrackIt's default values, which can be found at the above URL.

2.1.6 | Materials and apparatus

Stimuli were presented on a touchscreen laptop with physical dimensions 34.2 cm × 19.1 cm and pixel dimensions 1920 × 1080 (approximately 40° × 22° of visual field). Participants were seated at a desk facing the screen with their heads about 0.5 m away from the screen. While participants performed TrackIt, their gaze was recorded at 60 Hz using an SMI RED-250 mobile eye-tracker (SMI, 2009). Eye-tracker calibration and validation were performed automatically by SMI's iView X software, using default settings. These procedures each required the participant to fixate for at least 400 ms on each of a series of fixation points distributed over the display. A custom Python script (available at https://osf.io/vqjgs/) was then used to collect eye-tracking data synchronized with TrackIt.

2.1.7 | Data preprocessing

As described in Kim et al. (2020), we used a hidden Markov model (HMM) to infer the object to which a participant was attending at frame of each TrackIt trial. This algorithm outputs a high-frequency (60 Hz) temporal sequence of objects (from among the seven TrackIt objects) that the participant is estimated to be tracking. Kim et al. (2020) validated their HMM algorithm by showing that it agreed with judgements made by trained human coders and was able to identify attentional transitions (pairs of consecutive frames between which the inferred objects of attention differ) with accuracy comparable to agreement between human coders. The HMM has a free parameter, $\sigma \in (0, \infty)$, indicating the variance of the participant's gaze around the center of the object they were tracking; we used $\sigma = 300$, the value found by Kim et al. (2020) to maximize agreement with human coders. Further details of the HMM algorithm can be found in the manuscript of Kim et al. (2020).

Additionally, because child eye-tracking data contain a large proportion of missing values (including frames during which participants looked away from the screen), Kim et al. (2020) first linearly interpolated gaze positions for short intervals of missing data (< 10 frames, ≈ 167 ms) and then discarded all data from eight children with more than 50% of eye-tracking data missing in more than half of their trials. This left data from 42 children (420 total trials), ages 3.5–6 years ($M = 4.65$ years, $SD = 0.71$ years), 17 male and 25 female, for analysis. Before discarding data, the total proportion of trials missing more than 50% of eye-tracking data was 15.5% and the proportion of such trials per participant was not significantly correlated with Age ($r(48) = -0.10$, $p = 0.51$); participants had an average of 8622.69 frames of data, of which an average of 1819.38 (21.1%) were missing. After discarding data, 7.2% of trials were missing more than 50% of eye-tracking data, and participants had an average of 8628.04 frames of data, of which an average of 1819.38 (14.1%) were missing.

2.1.8 | Operationalizations of selective sustained attention

First, we review Location and Memory measures, based on participant response at the end of the trial, that have been used in previous TrackIt studies. Our interest in these measures is primarily in understanding how they relate to the eye-tracking-based measures we propose in the next section.

2.1.9 | Location Accuracy

Location Accuracy denotes the proportion of trials on which a participant correctly indicated the final grid cell of the Target object. This is the standard measure of task performance used in TrackIt. Note that, if the participant is not following the Target, Location Accuracy has a chance value of 1/36 on a 6×6 grid (as in Experiment 1) and 1/16 on a 4×4 grid (as in Experiment 2).

2.1.10 | Memory Accuracy

After providing a location response, the participant was presented with four distinct objects, one of which was the Target in that trial and three of which were Distractors in that trial. Memory Accuracy denotes the proportion of trials on which the participant correctly identified the Target object. This measure is used to partially distinguish Location errors due to forgetting, or otherwise being unaware of the Target object, from Location errors due to other attention failures.

We now describe the two main eye-tracking measures through which we operationalized selective attention maintenance (Staying) and selective transitioning (Returning) behaviors, based on the attention sequence inferred by the HMM algorithm (Kim et al., 2020). The design of these measures is a central methodological contribution of this paper, to support the distinction of Staying and Returning behaviors within SSA.

2.1.11 | Returning

As our measure of Returning, denoted $\widehat{R}$, we used the proportion of transitions from Distractors that go to Target:

$$\widehat{R} := \frac{\text{Number of transitions from Distractors to Target}}{\text{Number of transitions from Distractors to Any Object}}. \qquad (1)$$

$\widehat{R}$ measures a participant's tendency, when attending to a Distractor, to preferentially transition to the Target more than to other Distractors. $\widehat{R}$ takes values between 0 and 1, and, importantly, when a participant's attention is *not* selective to the Target object, $\widehat{R}$ has a "chance" expected value of 1/6 (since, from a given Distractor, the participant can transition to six other possible objects, one of which is the Target). This chance value is invariant to parameters of the HMM, eye-tracker noise, and other environmental or task factors, making $\widehat{R}$ a relatively pure measure of endogenous preference for transitioning to Target. When computing $\widehat{R}$, we omitted all frames during which gaze data were missing (including frames during which participants looked away from the display screen); hence, a transition was counted if the object of attention differed before and after the frames of missing gaze data.

2.1.12 | Staying

As our measure of Staying, denoted $\widehat{S}$, we used the average duration of continuous periods on Target minus the average duration of continuous periods on Distractors:

$$\begin{aligned}\widehat{S} :=\ & \text{avg(duration of periods on Target)} \\ & -\text{avg(duration of periods on a single Distractor)}. \qquad (2)\end{aligned}$$

This measures a participant's tendency to spend longer durations tracking the Target than tracking Distractors. Thus, when a participant's attention is *not* selective to the Target object, $\widehat{S}$ has a "chance" expected value of 0 s. This chance value is invariant to parameters of the HMM, eye-tracker noise, and other environmental or task factors, making $\widehat{S}$ a relatively pure measure of endogenous preference for maintaining attention on the Target. When computing $\widehat{S}$, periods of attention before and after periods of missing gaze data were treated as two separate periods of attention, regardless of whether the object of attention was the same during the two periods.

2.2 | Results

2.2.1 | Comparison with chance values

We first summarize participant performance in all behavioral and eye-tracking measures and compare each measure to its chance value. Ideally, participants should understand and be engaged with the task, but the task should also be challenging enough to observe distractions and prevent ceiling effects in performance; that is, mean Location Accuracy should be above chance but also be well below its maximum value of 1. Additionally, if SSA is endogenously supported through selective attention maintenance, then $\widehat{S}$ should be above chance, and, if SSA is endogenously supported through selective transitioning, then $\widehat{R}$ should be above chance. As shown in Table 1, participants performed above chance according to all four measures, suggesting that participants understood and were engaged in the task, and that SSA was endogenously supported through both Staying and Returning. The mean Location Accuracy of 40% suggests that overall task difficulty was well-calibrated to avoid both ceiling and floor effects.

2.2.2 | Correlations between measures

We next examined relationships between the four performance measures, as well as participant age. Our hypotheses were as follows. First, based on prior work in the TrackIt paradigm (Fisher et al., 2013; Kim et al., 2017), we expected that Location Accuracy would increase with Age, but that this improvement would not be explained by improvements in Memory Accuracy over Age. Second, assuming that TrackIt performance is supported by both Staying and Returning, we expected that both $\widehat{R}$ and $\widehat{S}$ would positively correlate with Location Accuracy. Third, given the lack of prior research into the distinct roles of selective attention maintenance and transitioning in the development of SSA, the relationships of greatest interest to us were those between $\widehat{R}$, $\widehat{S}$, and Age. In particular, we expected at least one of $\widehat{R}$ and $\widehat{S}$ to increase with Age, but it was not clear which, if not both of these, would increase with Age. Finally, although we expected both Staying and Returning to support TrackIt performance, it was not clear whether they would share underlying cognitive mechanisms, and we were, therefore, agnostic as to whether $\widehat{R}$ and $\widehat{S}$ would be directly related.

Table 2 shows Pearson correlation coefficients between all four performance measures and participant age, along with corresponding significance levels. All variables were positively correlated within the sample, but strength and statistical significance of these correlations varied.

TABLE 1 Univariate statistics for distributions (across 42 participants) of each of the four performance measures.

Measure	Chance value	Mean (95% CI)	Std. Dev.	$t(41)$-value	p-value
Location Accuracy	$1/36 = 0.02\overline{77}$	0.40 (0.31, 0.49)	0.29	8.32	<0.001***
Memory Accuracy	$1/4 = 0.25$	0.86 (0.80, 0.91)	0.18	21.96	<0.001***
$\hat{R}$	$1/6 = 0.1\overline{66}$	0.47 (0.41, 0.53)	0.20	9.83	<0.001***
$\hat{S}$	0	1.03s (0.77s, 1.29s)	0.86s	7.76	<0.001***

t-values and p-values are according to a 1-sample t-test for the null hypotheses that the means of the measures are equal to their chance values.
***$p < 0.001$.

TABLE 2 Matrix of Pearson correlations between age and performance measures used in this study. Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).

	Age	Location Accuracy	Memory Accuracy	$\hat{R}$	$\hat{S}$
Age	–	0.532***	0.263	0.544***	0.289
Location Accuracy	–	–	0.285	0.647***	0.496***
Memory Accuracy	–	–	–	0.336*	0.313*
$\hat{R}$	–	–	–	–	0.468**
$\hat{S}$	–	–	–	–	–

Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).
*$p < 0.05$, **$p < 0.005$, ***$p < 0.001$.

We first note that Location Accuracy improved strongly with Age. On the other hand, Memory Accuracy was not significantly correlated with Age or Location Accuracy. Since Memory Accuracy is close to ceiling and much higher than Location Accuracy (see Table 1), we conclude that children are consistently able to encode and maintain the identity of the Target object, and that improvement in this ability fails to explain children's improving ability to perform TrackIt with age. These results are all consistent with prior research using TrackIt in this age range (Fisher et al., 2013).

Both of our novel eye-tracking performance measures, $\hat{R}$ and $\hat{S}$, were strongly correlated with Location Accuracy, consistent with our expectation that Returning and Staying behaviors both support TrackIt performance. Additionally, $\hat{R}$ was strongly correlated with Age, which motivated us to perform further mediation analysis, described below, to investigate whether improvements in Returning explain improvements in TrackIt performance over age. On the other hand, $\hat{S}$ was not significantly correlated with Age. We discuss this key finding further in the Discussion. Figure 2 visualizes the main relationships between the two eye-tracking performance measures and both Location Accuracy and Age.

$\hat{R}$ and $\hat{S}$ were also strongly correlated. Since $\hat{S}$ was not significantly correlated with Age, this suggests that some interindividual differences supporting both Staying and Returning behaviors may not be fully explained by age. Finally, we note that Memory Accuracy was correlated with both $\hat{R}$ and $\hat{S}$, although we did not have hypotheses regarding these relationships.

2.2.3 | Mediation analysis

Since $\hat{R}$ was significantly correlated with both Age and Location Accuracy, we next used mediation analysis to study the extent to which

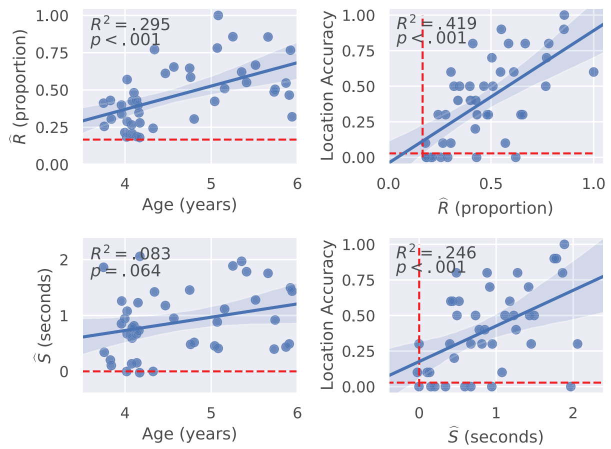

FIGURE 2 Left panels show linear regressions of $\hat{R}$ (Returning) and $\hat{S}$ (Staying) over Age. Right panels show linear regressions of Location Accuracy over $\hat{R}$ and $\hat{S}$. Shaded regions indicate bootstrapped 95% confidence bands. Dotted red lines indicate chance values for each performance measure.

improvements in Returning might explain improvements in Location Accuracy over age. We implemented mediation analysis according to the bootstrap procedure described in Preacher and Hayes (2008) with 10,000 bootstrap samples. As illustrated in Figure 3, the standardized indirect effect of Age on Location Accuracy mediated by $\hat{R}$ was 0.31, with 95% confidence intervals [0.16, 0.49]. The same bootstrap procedure yielded a p-value of $p < 0.001$ for the null hypothesis of 0 indirect effect. This suggests that Returning may play a strong mediating role between Age and Location Accuracy, explaining an estimated 59% (95% CI (28%, 96%)) of their common variation.

Since $\hat{S}$ was significantly correlated with both $\hat{R}$ and Location Accuracy, we also ran the above mediation analysis including $\hat{S}$ as a

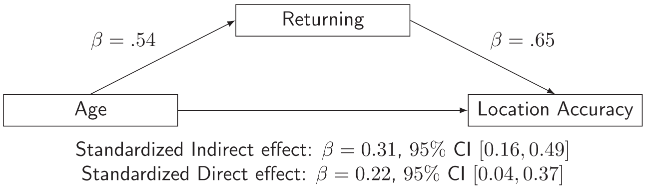

FIGURE 3 Standardized regression coefficients for the relationship between Age and Location Accuracy, as mediated by Returning. Confidence intervals for indirect and direct effects were computed based on a bootstrap procedure with 10,000 bootstrap samples, as described by Preacher and Hayes (2008).

control variable. Results were very similar, with a standardized indirect effect of Age on Location Accuracy (mediated by $\hat{R}$) of 0.18, with 95% confidence intervals $[0.13, 0.48]$. The same bootstrap procedure yielded a *p*-value of $p = 0.004$ for the null hypothesis of 0 indirect effect, and the estimated proportion of variation was 42% (95% CI (12%, 100%)).

2.2.4 | Differential development of returning and staying over Age

Above, we found that $\hat{R}$, but not $\hat{S}$, increased significantly over age. To compare the relative contributions of $\hat{R}$ and $\hat{S}$ in the development of SSA, we next investigated whether Age *differentially* affects Staying and Returning; that is, whether $\hat{R}$ increases *more* over Age than $\hat{S}$. This would strengthen our above finding that $\hat{R}$ increases with Age. We, thus, tested whether the correlation between $\hat{R}$ and Age was significantly greater than that between $\hat{S}$ and Age, using a test for comparison of correlations from dependent samples (Eid et al., 2011, p. 548). As hypothesized, the correlation between $\hat{R}$ and Age was significantly greater than that between $\hat{S}$ and Age ($z = 1.769$, $p = 0.038$), suggesting that Returning improves more with age than does Staying.

2.2.5 | Age-related changes in Average Tracking Duration

Although we did not find significant age-related increases in the *selectivity* of continuous attention maintenance (Staying), prior literature suggests that duration of attention ("attention span") continues to increase between the ages of 3.5 and 6 years (Moyer & Gilmer, 1955; Ruff & Rothbart, 2001). To investigate whether this development can be measured in TrackIt, we investigated whether the *Average Tracking Duration*, the average duration a participant spent on any object (Target or Distractor) changed with Age. Linear regression found a significant increase in Average Tracking Duration over Age ($\beta = 0.933$, $t(40) = 2.044$, $p = 0.048$). We caution that, unlike Staying, which measures the difference in participant behavior towards the Target and Distractors, Average Tracking Duration is not as carefully controlled a measure; it does not have a clear chance value to compare with and is likely more sensitive to missing data, eye-tracker calibration quality, and so forth.

2.2.6 | The effect of distance on Transitioning

Finally, we investigated exogenous influences on Staying and Returning. During each trial, only the positions of the objects changed over time, while features such as object color, shape, and size were fixed. Therefore, we reasoned that the primary exogenous influence on *when* participants transitioned between objects would be the objects' positions. Specifically, in a post hoc analysis motivated by prior research on bottom-up factors influencing visual attention (Desimone & Duncan, 1995; Desimone, 1998), we considered the distance of each object from the center of the participant's visual field to be a measure of the bottom-up salience of the object as it varied over the course of a trial. Because the measured gaze position on any particular frame can be quite noisy, we used the distance between the *center of the source object* and the center of the destination object, rather than the distance between the *gaze point* and the center of the destination object. Specifically, for each unattended object during each frame of the experiment, we recorded its distance from the current object of attention ("interobject distance"), as well as a binary indicator indicating whether the participant transitioned to that object on the next frame. If increasing object salience towards the center of the participant's visual field supports transitioning to an object, then we would expect the frequency of transitions to decrease as a function of interobject distance. Additionally, if the process of becoming distracted is more sensitive to bottom-up influences than the process of Returning, then we would expect this decreasing relationship to be stronger for Target-to-Distractor transitions than for Distractors-to-Target transitions.

To quantify the frequency of each type of transition, we tabulated all of the *possible* transitions (i.e., 6 per frame, one for each possible destination object from the current object of attention). We coded those transitions that actually occurred (i.e., when the participant transitioned to that destination object on the next frame) with the value 1, and coded all other transitions with the value 0. Figure 4 illustrates this computation visually. We then regressed this indicator variable over the distance between the two objects, separately for each transitions type (Target-to-Distractor, Distractor-to-Distractor, or Distractor-to-Target).

Figure 5 shows the nonlinear effect of inter-object distance on the frequency of transitioning to a given destination object on a given frame, separated by the type of possible transition (Target-to-Distractor, Distractor-to-Distractor, or Distractor-to-Target). Table 3 shows the mean (over participants) correlation between each kind of transition probability and inter-object distance. Our main observation was that the probability of Target-to-Distractor transitions increased with decreasing interobject distance.

Additionally, the probability of Distractor-to-Target transitions increased with increasing interobject distance. In Figure 5, the probability of Target-to-Distractor transitions also showed a small local increase between the distances of 5° and 8°. We believe that these effects may be artifacts of the data analysis process because short-distance transitions are less likely to be identified by the HMM,

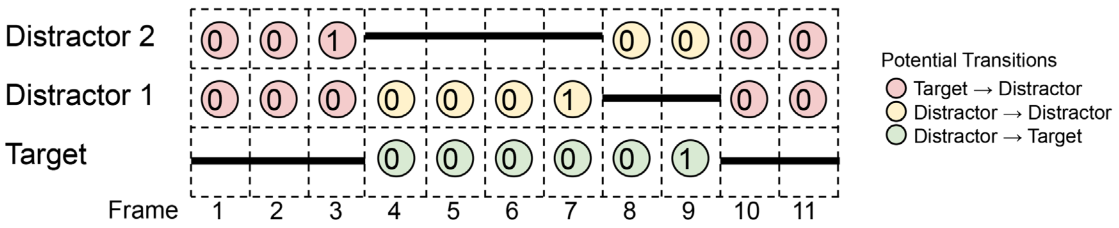

FIGURE 4 Illustration of data used in the "The Effect of Distance on Transitioning" analysis, in a simplified example with three objects over 11 frames. The dark black line indicates the object of attention (as inferred by the HMM) in each frame. Each circle corresponds to one possible transition after each frame, coded as 1 when the participant performed that transition and 0 otherwise. The circle's color indicates the type of transition (red for Target-to-Distractor, yellow for Distractor-to-Distractor, and green for Distractor-to-Target).

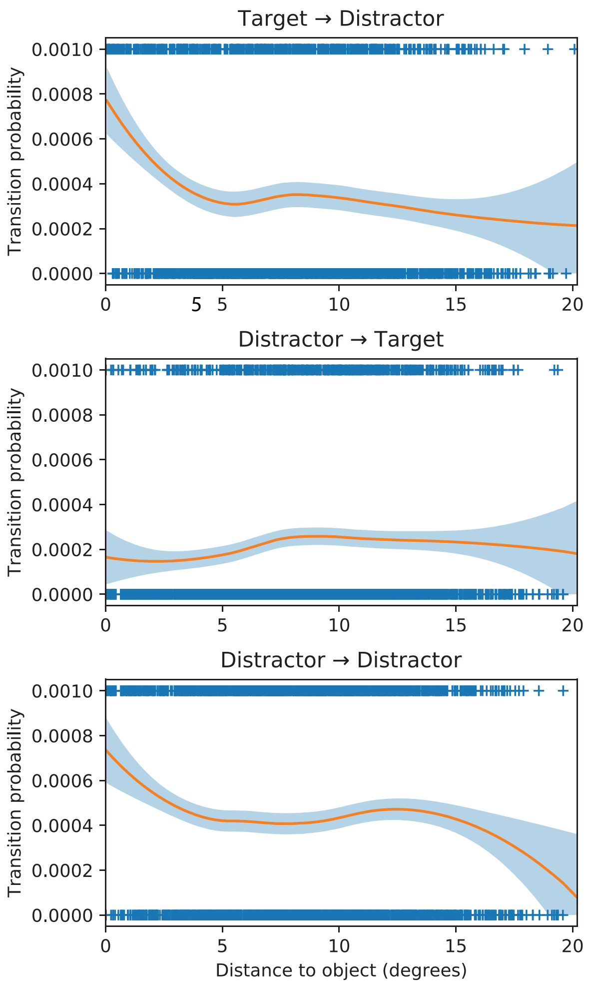

FIGURE 5 Nonparametric regression of empirical transition probability (i.e., probability of transitioning to a particular destination object given a particular source object) over distance between source and destination objects, for different source and destination object types. LOESS regression curves are plotted in orange and 95% confidence bands are shaded light blue. Blue crosses show a representative sample of object distances corresponding to transitions (coded as 0.001) and nontransitions (coded as 0). Note that, because our sampling rate was quite high (60 Hz), transitions occur on only a very small proportion (∼ 0.1%) of frames. Hence, nontransition frames were randomly down-sampled by a factor of 1000 for visualization.

resulting in a downward bias of estimated transition probabilities for small interobject distances. This bias is independent of the types of the source and destination objects, and should, therefore, not affect our main finding, namely that the distance from the center of the visual field decreases the likelihood of Target-to-Distractor transitions more than Distractor-to-Target transitions.

A potential alternative cause of the observed decrease in transition probabilities at short distances could be an inhibitory effect of proximity on the salience of unattended objects, such as surround suppression Allman et al. (1985); Cavanaugh et al. (2002); indeed, top-down attentional effects on surround suppression have been identified even in early stages of visual processing (Moran & Desimone, 1985; Sundberg et al., 2009). However, for such an effect to explain the main differential effect of interobject distance on Distractor-to-Target and Target-to-Distractor transitions, the top-down attentional suppressive effect would have to be *stronger* when a participant is attending to a Distractor than when they are attending to the Target. This would be contrary to established models of top-down effects on visual attention (Desimone & Duncan, 1995; Desimone, 1998), which would predict greater top-down effects when on Target.

To conclude, we found that participants were more likely to transition from Target to a Distractor when the Distractor was nearer to the Target, while such an influence was not observed for transitions from a Distractor to Target. Before further discussing our findings, we present a second experiment that strengthens the generalizability of our findings and confirms our post hoc analyses on the effect of distance on transitions.

3 | EXPERIMENT 2

Some analyses presented in Experiment 1, such as the analysis of the effect of interobject distance on transition probabilities, were post hoc. To perform confirmatory versions of these analyses and strengthen external validity of the findings reported in Experiment 1, in Experiment 2, we conduct the full suite of analyses conducted in Experiment 1 on an additional dataset of behavioral and eye-tracking data in the TrackIt task. Because this dataset was collected for a different purpose, it presents several design differences relative to Experiment 1, providing a test of generalizability of the findings of Experiment 1.

TABLE 3 Distributional statistics (across 42 participants) of the correlation (across all frames within each participant) between interobject distance and each type of transition probability (with transitions coded as 1 and nontransitions coded as 0).

Transition type	Mean correlation	Standard deviation	t(41)-value	p-value
Target → Distractor	−0.10	0.13	−4.83	<0.001***
Distractor → Target	0.06	0.10	3.73	<0.001***
Distractor → Distractor	−0.007	0.10	−0.3	0.76

p-values are according to 1-sample t-tests for the null hypotheses that the mean correlations are 0.
***$p < 0.001$.

3.1 | Methods

Full details of this experiment can be found in Vales et al. (2021). Here we describe the main features of this dataset, focusing on design differences relative to Experiment 1.

3.1.1 | Participants

We analyzed data from one of two experimental conditions (described below) reported in Vales et al. (2021), including data from 36 participants, aged 4–6 years ($M = 4.57$ years, $SD = 0.68$ years), 19 male and 17 female, that provided usable eye-tracking data (after the same preprocessing steps described in Experiment 1). Race and ethnicity information for the sample reported by parents or guardians was as follows: 50% White, 5.6% Black or African American, 8.3% East Asian or Asian American, 2.8% Middle Eastern or Arab American, and 33.3% unreported.

3.1.2 | Materials, design, procedure, and data preprocessing

Stimulus presentation apparatus, response collection, and eye-tracking collection and preprocessing were analogous to Experiment 1. Participants were randomly assigned to one of two conditions designed to examine the role of verbal labels and shape familiarity on SSA. In the Labeled condition, the Target shape was explicitly labeled before the trial started (e.g., "Follow the triangle") and in the Unlabeled condition it was not (e.g., "Follow the shape"); in both conditions, the Target was indicated visually by a red circle before trial onset (as in Figure 1). Since evaluating the effect of labels is beyond the scope of this paper, we analyzed only data from the Unlabeled condition in the main paper, and we report results on the Labeled condition, which were largely similar to those on the Labeled condition, in Appendix A.3. In addition to using prerecorded audio instructions (e.g., "Follow the shape"), the stimuli and design differed from Experiment 1 in five ways. First, all stimuli were colored blue, so that shape was the sole distinguishing feature among stimuli in a trial. Second, participants completed 12 trials (excluding an initial practice trial, not analyzed). Third, only four Distractor were presented, for a total of five TrackIt objects in each trial. Fourth, the grid size was 4×4 (with each grid cell the same size as cells in the 6×6 grid of Experiment 1). Finally, only six different shapes (circle, triangle, square, pentagon, diamond, and oval) were used. Objects were classified as "High Familiarity" (circle, triangle, square) or "Low Familiarity" (pentagon, diamond, oval), based on their name's frequency in the CHILDES corpus (Bååth, 2010; MacWhinney, 2000). Each object served as Target on 2 (randomly ordered) trials and was among the four shapes from which Distractors were sampled on the remaining 10 trials, with the constraint that each trial included two High Familiarity Distractors and two Low Familiarity Distractors. We note that none of these design differences caused us to alter our hypotheses from Experiment 1.

3.2 | Results

All analyses were identical to those performed in Experiment 1. For brevity, we simply state the results; refer to Section 2.2 of Experiment 1 for motivation of each analysis.

3.2.1 | Comparison with chance values

As shown in Table 4, participants performed above chance according to all four measures, suggesting that participants understood and were engaged in the task, and that SSA was endogenously supported through both Staying and Returning. These findings all replicate those in Table 1 of Experiment 1.

3.2.2 | Correlations between measures

Table 5 shows Pearson correlation coefficients between all four performance measures, as well as participant age, along with corresponding significance levels.

Location Accuracy improved strongly with Age, whereas Memory Accuracy was not significantly correlated with Age. Since Memory Accuracy is close to ceiling and much higher than Location Accuracy (see Table 4), we conclude that children are consistently able to encode and maintain the identity of the Target object, and that improvement in this ability fails to explain children's improving TrackIt performance over age. Both of our novel eye-tracking performance measures, $\widehat{R}$ and $\widehat{S}$, correlated positively with Location Accuracy, consistent with our expectation that Staying and Returning behaviors both support TrackIt performance. Additionally, $\widehat{R}$ increased significantly

TABLE 4 Basic univariate statistics for distributions (across 36 participants) of each of the four performance measures in Experiment 2.

Measure	Chance value	Mean (95% CI)	Std. Dev.	t(35)-value	p-value
Location Accuracy	$1/16 = 0.0625$	0.45 (0.37, 0.54)	0.28	9.71	<0.001***
Memory Accuracy	$1/4 = 0.25$	0.77 (0.70, 0.85)	0.26	13.13	<0.001***
$\hat{R}$	$1/4 = 0.25$	0.51 (0.45, 0.57)	0.19	11.61	<0.001***
$\hat{S}$	0	1.13s (0.88s, 1.37s)	0.80s	9.85	<0.001***

p-values are according to 1-sample t-tests for the null hypotheses that the means of the measures are equal to their chance values.
***$p < 0.001$.

TABLE 5 Matrix of Pearson correlations between age and performance measures in Experiment 2.

	Age	Location Accuracy	Memory Accuracy	$\hat{R}$	$\hat{S}$
Age	–	0.437**	0.292	0.385*	0.061
Location Accuracy	–	–	0.669***	0.659***	0.383*
Memory Accuracy	–	–	–	0.539***	0.271
$\hat{R}$	–	–	–	–	0.378*
$\hat{S}$	–	–	–	–	–

Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).
*$p < 0.05$, **$p < 0.005$, ***$p < 0.001$.

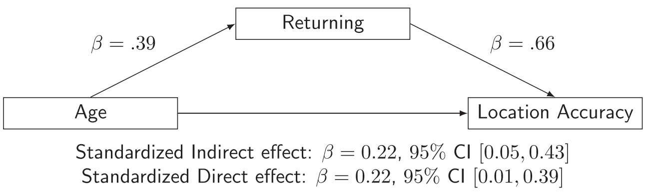

FIGURE 6 Standardized regression coefficients for the relationship between Age and Location Accuracy, as mediated by Returning, in Experiment 2. Confidence intervals for indirect and direct effects were computed based on a bootstrap procedure with 10,000 bootstrap samples, as described by Preacher and Hayes (2008).

with Age, whereas $\hat{S}$ did not. Finally, $\hat{R}$ and $\hat{S}$ were again significantly correlated.

These findings all replicate those of Table 2 in Experiment 1. The only notable difference (in terms of statistical significance) from Table 2 was that, here, Memory Accuracy was significantly correlated with Location Accuracy. Since, in this experiment, all objects had the same color, this may have been because encoding the distinct identity of the Target object was more difficult in this experiment and, therefore, played a bigger role in Location Accuracy performance than in Experiment 1; indeed, Memory Check performance ($M = 0.77, SD = 0.26$) was lower than in Experiment 1, ($M = 0.86, SD = 0.18$), although this difference was not statistically significant ($t(78) = 1.80, p = 0.076$).

3.2.3 | Mediation analysis

As illustrated in Figure 6, mediation analysis replicated results of Experiment 1, with the indirect effect explaining an estimated 51% (95% CI: (14%, 100%)) of common variation between Age and Location Accuracy. This mediation effect was statistically significant ($p = 0.003$) according to the bootstrap procedure described in Preacher and Hayes (2008).

When controlling for $\hat{S}$, the results of the mediation analysis were similar: the estimated indirect effect was 0.18 (95% CI: (0.03, 0.37), the bootstrap p-value was $p = 0.011$, and the estimated proportion of mediation was 44% (95% CI: (6%, 100%)).

3.2.4 | Differential development of Returning and Staying over Age

As in Experiment 1, the correlation between $\hat{R}$ and Age was significantly greater than that between $\hat{S}$ and Age ($z = 1.746$, $p = 0.040$), suggesting that Returning develops more over this age range than does Staying.

3.2.5 | Age-related changes in Average Tracking Duration

As in Experiment 1, participants' Average Tracking Duration increased significantly over Age ($\beta = 2.90, t(34) = 2.77, p = 0.008$).

3.2.6 | The effect of distance on Transitioning

The effect of distance on each type of transition probability, illustrated in Figure 7, was qualitatively similar to that in the Experiment 1. Linear regression of transition frequencies over interobject distance, detailed

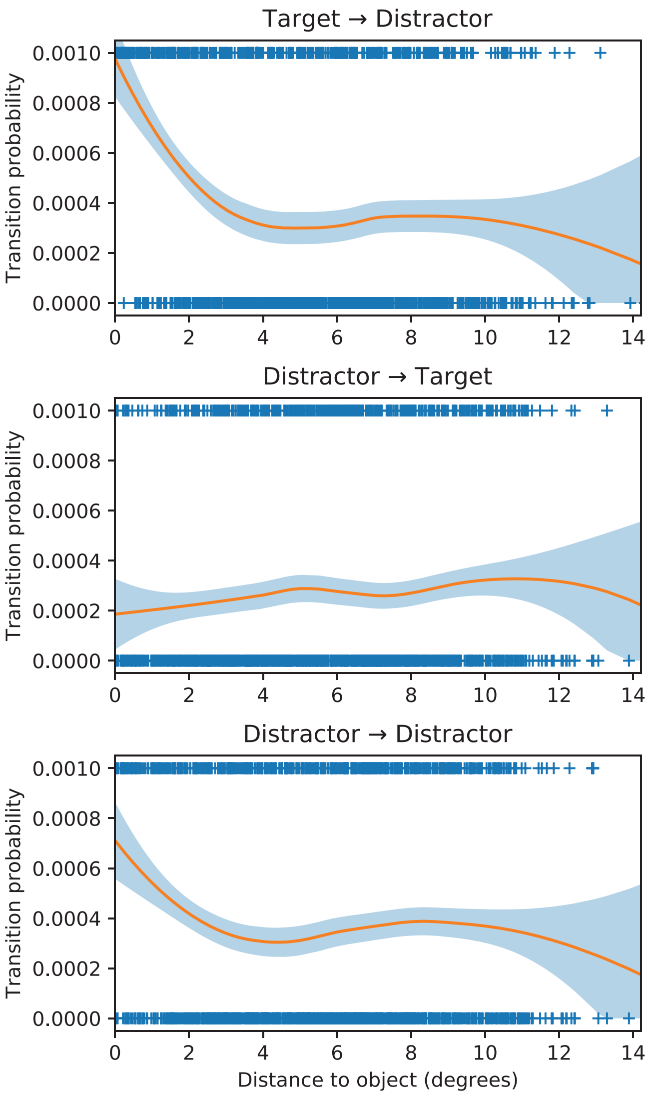

FIGURE 7 Nonparametric regression of empirical transition probability (i.e., probability of transitioning to a particular destination object given a particular source object) over distance between source and destination objects, for different source and destination object types, in Experiment 2. LOESS regression curves (orange) and 95% confidence bands (shaded blue) were computed by Python's SciPy package Virtanen et al. (2020) using default parameters. Blue crosses show a representative sample of object distances corresponding to transitions (coded as 0.001) and nontransitions (coded as 0). Nontransition frames were randomly downsampled by a factor of 1000 for visualization.

in Table 6, again showed that the probability of Target-to-Distractor transitions decreases with interobject distance, while the probability of Distractor-to-Target transitions increases with interobject distance, and the probability of Distractor-to-Distractor transitions does not significantly change with interobject distance.

3.2.7 | Summary of Experiment 2

Experiment 2 successfully replicated all of the main findings of Experiment 1. The only notable difference between the results of the two experiments was that, in Experiment 2, Memory and Location Accuracies were significantly correlated, whereas, in Experiment 1, they were not. However, since Memory Accuracy did not improve significantly with age, this difference does not affect our developmental conclusions. As noted previously, findings for the Labeled condition, reported in Appendix A.3 largely mirrored those reported for the Labeled condition in this section.

4 | GENERAL DISCUSSION

4.1 | Summary of findings

Our main findings are as follows. First, $\widehat{R}$ and $\widehat{S}$, respectively, measuring Returning (selective transitioning) and Staying (selective attention maintenance), were both significantly larger than their chance values under the null hypothesis of nonselectivity. While $\widehat{R}$ and $\widehat{S}$ both correlated strongly with TrackIt task performance, only $\widehat{R}$ improved significantly with Age; additionally, $\widehat{R}$ increased significantly more with Age than did $\widehat{S}$. Mediation analysis suggested that these improvements in Returning explain a significant proportion of age-related improvements in TrackIt performance, even when controlling for changes in $\widehat{S}$. Meanwhile, *nonselective* attention maintenance, measured by Average Tracking Duration, also increased significantly with age. Additionally, transitions from the Target to a Distractor were more likely when the Distractor was closer in distance to the Target, whereas this effect was not observed for transitions from Distractors to the Target. As we discuss below, all of these results are consistent with the idea that Returning, more so than Staying, relies on endogenous control of attention, which exhibits a protracted developmental time course. These findings were all replicated in confirmatory analysis of an independent dataset in Experiment 2.

4.2 | Endogenous influences supporting target selectivity

A key component of SSA on a task is selectivity, the preferential allocation of attentional resources to task-relevant loci, such as the Target object in TrackIt. As discussed in the Introduction, selectivity for the Target could manifest as transitioning from Distractors to the Target *more frequently* than to other Distractors (increased Returning), or as maintaining attention *longer* on the Target than on Distractors (increased Staying), or as both.

The finding that mean $\widehat{R}$ is greater than its chance value shows that when attending to a Distractor, participants preferentially return to the Target over transitioning to other Distractors. Since, when a participant is tracking a Distractor, there is no exogenous support for selectively transitioning to the Target, this preference implies an endogenous selective influence on attentional transitions. Note that this does not exclude possible simultaneous contribution of exogenous factors (see, e.g., our discussion of possible "rescue saccades," along with further discussion of connections to the Multiple Object Tracking task, in Appendix A.5) in Returning.

TABLE 6 Distributional statistics (across 36 participants in Experiment 2) of the correlation (across all frames within each participant) between interobject distance and each type of transition (with transitions coded as 1 and nontransitions coded as 0). *p*-values are according to 1-sample *t*-tests for the null hypotheses that the mean correlations are 0.

Transition type	Mean correlation	Standard deviation	$t(35)$-value	p-value
Target → Distractor	−0.15	0.16	−6.183	<0.001***
Distractor → Target	0.09	0.15	3.64	<0.001***
Distractor → Distractor	−0.02	0.15	−1.11	0.28

***$p < 0.001$.

Since $\hat{S}$ was defined as the average duration on the Target minus that on Distractors, the finding that mean $\hat{S}$ is greater than its chance value of 0 seconds (i.e., the null hypothesis that the average duration on the Target is equal to the average duration on a Distractor) shows that participants preferentially stay on the Target for longer than on Distractors. While this might suggest an endogenous selective influence on attention maintenance, the interpretation of this result is slightly less clear than for $\hat{R}$. Because the Target is indicated by a red circle before the trial begins, participants typically begin the trial looking at the Target, which may increase $\hat{S}$ artificially even in the absence of endogenous influences. In the extreme case, it is possible that visual attention is maintained on the present object of attention even in the absence of any endogenous influence, and thus participants might appear to maintain attention preferentially to the Target simply because they begin trial on Target.

We investigated the possibility of this bias through post hoc analysis in which we removed the first period of attention maintenance on an object within each trial and reexamined whether $\hat{S}$ was still significantly greater than its chance value of 0 s (see Appendix A.1 for details this analysis). $\hat{S}$ was significantly above chance even after this initial period of attention. This further supports the existence of an endogenous selective influence on attention maintenance that is not explained simply by participants beginning trial on Target.

To summarize, our data suggest that top-down selectivity for the Target manifests both as transitioning from Distractors to the Target more frequently than to other Distractors (increasing Returning) and as maintaining attention longer on the Target than on Distractors (increasing Staying). Additionally, at least for relatively nearby objects, the likelihood of transitioning from the Target to a Distractor increased with proximity of that Distractor to the Target, while we did not observe this effect for transitions from Distractors to the Target. If this effect is driven by increased salience of Distractors nearing the center of the visual field (Desimone & Duncan, 1995; Desimone, 1998), these results are consistent with the hypothesis that, whereas the process of becoming distracted may be initiated by bottom-up influences, the process of returning attention to the task is more likely to be initiated by a top-down influence. As we discuss in Section 4.3, further work is needed to investigate this hypothesis.

4.3 | Development of Staying and Returning

We found that $\hat{R}$ increased with age, suggesting that attention transitions become more selective with development. Additionally, mediation analysis suggested that this increase in selectivity of transitioning may explain a large proportion of improvement in TrackIt performance (measured by Location Accuracy) over age. Since Memory Accuracy was already quite high in the youngest participants and did not increase significantly with age, these improvements in Returning are unlikely to be explained by improvements in encoding the Target object's identity, although they could be supported by improvements in other aspects of working memory (e.g., goal maintenance). Meanwhile, any increases in $\hat{S}$ over age were not statistically significant, and, as correlations, were significantly weaker than increases in $\hat{R}$ over age. This suggests that the improvements in TrackIt performance with development between the ages of 3.5 and 6 years may stem more from increased selectivity of transitions than from increased selectivity of attention maintenance. Moreover, the increase in Returning appears unlikely to be a mere artifact of general factors, such as increases in task engagement over age, that would likely lead to increases in both Staying and Returning over Age.

4.3.1 | Cognitive mechanisms underlying Returning

Assuming that the improvements we observe in Returning are supported by the development of endogenous control of attention, our results are consistent with previous work showing development of endogenous attentional control and reduced dependence on exogenous support of attention in TrackIt performance in this age range (Fisher et al., 2013). Two particular endogenous processes that may support Returning improvements are working memory and inhibition. Both working memory and inhibition are thought to improve significantly in this age range (Diamond, 2002), likely supported by maturation of the prefrontal cortex (PFC) during these years (Casey et al., 2000; Diamond et al., 2004), and may be crucial for aspects of Returning such as maintaining a representation of the Target while following a Distractor, terminating attention to a Distractor based on representation of the task goal, or inhibiting Distractors while searching for or reorienting attention to the Target (Colombo & Cheatham, 2006; Kane & Engle, 2002; Ruff & Rothbart, 2001).

Since, when a participant is tracking a Distractor, there is no exogenous support for selectively transitioning to the Target, Returning should require some degree of working memory supporting encoding and maintaining internal representations of the Target and task goal. The relatively small improvement in Memory Check Accuracy over age suggests that improvements in encoding and maintaining the identity of the Target object alone are insufficient to explain improvements in Returning and in overall task performance. In particular, since Memory Check Accuracy was close to ceiling (and much higher than Location Accuracy), the ability to encode the Target (and its distinct task-relevance as compared to Distractors) appears to already be sufficiently developed in 3-year-olds. On the other hand, it remains possible that children's ability to encode and maintain the *current task goal* ("active goal maintenance"), which is hypothesized to be a key element of executive function (Friedman et al., 2008; Kane & Engle, 2002), improves. This also appears consistent with findings that children become more able to store an internal representation that conflicts with the externally attended stimulus (e.g., in the Day-Night task or theory of mind), as one must maintain an internal representation of the task goal and Target object even while following a Distractor (Diamond, 2002, 2006). Note that, since the memory check presents only four static, well-spaced objects, it is possible that the demands of the memory check on accurately encoding the Target identity are less than those in the main TrackIt task, thus not reflecting the same degree of Target encoding ability necessary in the main TrackIt task. To narrow this possibility, it may be desirable in future work to use a more difficult memory check that more closely reflects conditions in the main TrackIt task (e.g, displaying all seven objects from the trial in randomized or dynamic spatial configurations, as in Fisher et al. (2013)).

Returning may also rely on inhibitory control to prevent transitioning to different Distractor objects. Because Returning requires behaving based on a joint function of the stimulus and task goals, this would likely require what Munakata et al. (2011) refer to as "indirect competitive" inhibition within neocortical regions, in which prefrontal regions selectively enhance task-relevant representations, which in turn inhibit task-irrelevant representations through lateral competitive mechanisms. This is distinguished from what Munakata et al. (2011) refer to as "directed global" inhibition of subcortical regions, where prefrontal regions select which functions of subcortical regions should be suppressed, resulting in a more general inhibitory state that is not selective to the stimulus. Specifically, since Returning is selective to the Target, we believe that the former may be the relevant type of inhibition. The improvements we observed in Returning over age are consistent with prior findings that indirect competitive inhibition improves between the ages of 3 and 5 years Diamond (2002), in contrast to directed global inhibition, which is believed to mature earlier in development (Munakata et al., 2011).

4.3.2 | Cognitive mechanisms underlying Staying

Our results suggest that, while children exhibit significantly above chance Staying ability, Staying does not develop significantly or as much as Returning in this age range. This may suggest that the components of endogenous control that underlie Staying are already sufficiently developed by this age range. Endogenous control could support Staying through at least two mechanisms, by either inhibiting or exciting attentional transitions from the current object, corresponding to the Target and Distractor durations in terms of which Staying is defined (Equation 2). In the "inhibitory" mechanism, endogenous control might facilitate maintenance of attention on the Target by inhibiting transitions when the participant is attending to the Target, thereby increasing the duration of attending to the Target (the first term of Equation 2). In the "excitatory" mechanism, endogenous control might facilitate disengagement from the current object when the participant is attending to a Distractor, thereby shortening the duration of attending to a Distractor (the second term of Equation 2). Note that, because these mechanisms are agnostic to the *destination* of the transition, and only selective to the *source* of the transition (i.e., the current object of attention), we suggest these as potentially relevant mechanisms only for Staying, not for Returning.

Our finding that Average Tracking Duration increased significantly over age suggest an increase in the inhibition of attentional transitions over development between 3.5 and 6 years. However, since Staying did not increase significantly, the increased inhibition may be *nonselective*; that is, it may not depend on the task-relevance of the current object of attention.

Meanwhile, if we interpret the results of the distance analysis to suggest that transitions from Distractors to the Target are less sensitive to bottom-up salience than transitions from the Target to Distractors, then this may suggest the latter excitatory mechanism, wherein endogenous control initiates transitions from Distractors to the Target. This initiation, which requires first recognizing that one is distracted, might be supported by an endogenous supervisory performance monitoring system such as that described by Stuss et al. (1995) in their model of vigilant attentional control (Langner & Eickhoff, 2013; Shallice et al., 2008) Specifically, Stuss et al. (1995) outlined a set of four supervisory subprocesses that support vigilant attention: "(a) monitoring the activation level of the task schema, (b) (re)activating ("energizing") the task schema, (c) inhibiting conflicting schemata, and (d) monitoring performance (i.e., checking the appropriateness of behavioral outputs against the task goal)" (Langner & Eickhoff, 2013, p. 871). In particular, the frequency and success with which each of these subprocesses occurs would contribute directly to how quickly one is able to disengage from a Distractor, thereby shortening the average duration of periods on single Distractors and thus contributing to Staying.

4.4 | Limitations

Our measures of Staying and Returning are designed around the assumption that the participant is tracking the Target object. However, participants may, for reasons unrelated to failures of attention (e.g., forgetting which object is the Target), track other objects. In this case, effects we observed in Staying and Returning (e.g., improvement in Returning over age) might stem from changes in knowing which object to track, rather than from any changes in the processes of Staying and Returning themselves. Appendix A.4 presents supplementary analyses in which we reran the analyses of Experiment 1 using only the subset of trials on which participants correctly identified the Target object at the end of the trial (correct Memory Response). The results of this analyses analyses were nearly identical to those reported in Experiment 1, consistent with the idea that our findings are intrinsic to the Staying and Returning processes, rather than stemming from improvements in encoding the Target object.

Our analysis of the effect of interobject distance on the probabilities of different transition types was motivated by the assumption that the bottom-up salience of an object increases with its proximity to the center of the visual field. While research on early visual processing (Desimone & Duncan, 1995; Desimone, 1998) provides considerable evidence for this increased bottom-up salience, it is not certain whether this increased bottom-up salience is the mechanism by which interobject distance influences Target-to-Distractor transition probabilities. For example, another possibility is that participants overtly explore Distractors that are close to the Target because it is easy to do so while continuing to track the Target covertly. Appendix A.7 reports a post hoc investigation of whether the shape of an object, which should affect attentional transitions only through bottom-up mechanisms, had any effect on the probability of transitioning to that object. However, no such effect was found; this was unsurprising, since the shapes in this study were not designed to differ significantly in bottom-up salience. Future studies should clarify the relative effects of bottom-up factors on different types of attentional transitions by directly experimentally manipulating bottom-up salience of different objects; for example, TrackIt allows for many such manipulations, for instance by manipulating the shapes and colors of various objects, either between trials or at particular time points within a trial.

Another important technical limitation is that participants' true attentional states may not be perfectly reflected in the inferred attentional state sequence used as a starting point for data analysis in this study, due to both potential biases introduced by the HMM algorithm of Kim et al. (2020) and the imperfect coupling of attentional state and gaze behavior. To investigate possible biases introduced by the HMM algorithm, we replicated our results on a dataset of attentional state labels created by trained human coders using video reconstructions of the TrackIt trials overlaid with participant's gaze. These results, reported in Appendix A.2, replicated all of the main results reported in this paper, providing evidence for the validity of the attentional state inferred by the HMM algorithm.

On the other hand, gaze behavior itself may be an imperfect reflection of visual attentional state for at least two reasons. First, some spurious eye-movements may not reflect attending behavior. However, while unconscious object tracking (bottom-up eye-movements not reflecting conscious perception) have been observed in response to very specific artificial visual stimuli (e.g., small subthreshold visual manipulations Tavassoli & Ringach, 2010 or two conflicting stimuli, each presented to a different eye Spering et al., 2011), perceptual attention is believed to subserve smooth pursuit Kowler (1995), and the coupling between eye-movements and perception is believed to be strongest in cases of voluntary smooth pursuit Spering and Carrasco (2015), as we expect in TrackIt. Given this, we generally expect participants' eye-movements to be a function of attending behavior. Second, some attending behavior, namely covert attention, may not influence eye-movements. We attempted to minimize this possibility by explicitly instructing participants to follow the Target "with [their] eyes." Nevertheless, we are investigating whether children performing TrackIt utilize covert attention in follow-up work.

Finally, one question left open by our work concerns the nature of the task representation children maintain while they complete the TrackIt task. Further work is needed to understand the relative contributions of (a) verbal instructions or demonstration of the task and (b) the initial visual salience of the Target to the improvements in Returning over age. This could help distinguish the relative contributions of developments in visual working memory, active goal maintenance, and language understanding. For example, considering that verbal instructions have been shown to alter visual processing in young children Vales and Smith (2015, 2018), would we still observe improvements in Returning in the absence of verbal instructions or if the Target was identified verbally but not visually? Alternatively, if Staying and Returning are supported primarily by a visual representation of the Target, then modifications to the appearance of the Target during the trial (e.g., a color shift) would likely impair performance more than if they are supported by a linguistic shape representation (e.g., "follow the star"). While these questions about representation are outside the scope of the current investigations, the Track-It paradigm lends itself well to parametric variations in attentional loci that may be fruitfully used to investigate these questions.

4.5 | Applications

The ideas, results, and methods presented in this paper have connections to several areas of attention-related research and its applications; we conclude our discussion by noting a few of these. First, understanding children's ability to return attention to a task in the absence of exogenous cues may inform the design of supportive environments and materials for learning (Erickson et al., 2015; Eng et al., 2020; Fisher et al., 2014). For example, Eng et al. (2020) measured distractiveness of illustrated reading materials in terms of the number of gaze transitions towards task-irrelevant illustrations; however, such gaze transitions may be much more consequential for reading comprehension in younger children than in older children, due to the development of Returning, the ability to return to task-relevant information after distractions. Hence, the amount of extraneous (but potentially enriching) information that can be provided to children without sacrificing comprehension of domain-relevant information could be adjusted to age. Second, our distinction of Staying and Returning components of endogenous control over SSA may be relevant for understanding the cognitive underpinnings of mindfulness and other meditation practices, which often center around improving self-regulation of attention and often specifically include the practice of returning to a target locus once distracted (Jensen et al., 2012; Kabat-Zinn, 2003; Lutz et al., 2008). Finally, explicitly measuring and quantifying attentional transitions, as we begin to do in this paper, is an important step towards investigating the idea of a developmental trend along the exploration-exploitation trade-off in the guidance of attention, which has been raised in recent work (Blanco & Sloutsky, 2020; Deng & Sloutsky, 2016;

Gopnik, 2020; Plebanek & Sloutsky, 2017). In particular, whereas these papers have relied on indirect measures, such as participant recall after the task, direct measurement of attentional transitions could provide much more reliable measures of exploratory behavior.

5 | CONCLUSIONS

Extensive work has characterized children's ability to maintain focused attention on an assigned task over the course of development and related this ability to performance in school and later in life (Diamond, 2016; Fisher et al., 2013). By comparison, much less is known about how children allocate and regulate their attention when off task. This research provides evidence that (a) young children have an endogenous ability, "Returning," to return their attention to task after distractions, (b) Returning can be distinguished from selective maintenance of attention, "Staying," (c) Returning improves between the ages of 3.5 and 6 years, and (d) this improvement partly underlies previously identified improvements in SSA between these ages.

ACKNOWLEDGMENTS

We thank Melissa Pocsai, Oceann Stanley, and Tishya Girdhar for their help collecting data. We thank Priscilla Medor and Kristen Boyle for their help coding eye-tracking videos. We thank the children, parents, and teachers of the CMU Children's School, Amazing Scholars Academy Preschool, Beth Shalom Early Learning Center, and Glenn Avenue Preschool for making this work possible.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B150008 to Carnegie Mellon University, and by the U.S. National Science Foundation (NSF) through Grant BCS-1451706. The opinions expressed are those of the authors and do not represent the views of the Institute of Education Sciences, the U.S. Department of Education, or the NSF.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data are publicly available on the Open Science Framework (OSF). Data for Experiment 1 Kim et al. (2020) are available at https://osf.io/u8jbs. Data for Experiment 2 (Vales et al., 2021) are available at http://osf.io/kxbve. Python code for reproducing our analyses is available on GitHub at https://github.com/Jaeah/staying-and-returning.

ETHICS APPROVAL STATEMENT

All procedures were approved by the Carnegie Mellon University institutional review board. Informed consent was obtained from a parent or guardian of all participants included in the study.

ORCID

Jaeah Kim https://orcid.org/0000-0002-0462-086X
Anna V. Fisher https://orcid.org/0000-0002-6661-7922

REFERENCES

Akshoomoff, N. (2002). Selective attention and active engagement in young children. *Developmental Neuropsychology*, *22*(3), 625–642.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction-and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, *14*(2), 105–126.

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

Bae, G. Y., & Flombaum, J. I. (2012). Close encounters of the distracting kind: Identifying the cause of visual tracking errors. *Attention, Perception, & Psychophysics*, *74*(4), 703–715.

Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics*, *51*(2), 145–162.

Bennett, N., Jordan, J., Long, G., & Wade, B. (1976). *Teaching styles and pupil progress*. Open Books.

Blanco, N. J., & Sloutsky, V. M. (2020). Attentional mechanisms drive systematic exploration in young children. *Cognition*, *202*, 104327.

Brockhoff, A., Papenmeier, F., Wolf, K., Pfeiffer, T., Jahn, G., & Huff, M. (2016). Viewpoint matters: Exploring the involvement of reference frames in multiple object tracking from a developmental perspective. *Cognitive Development*, *37*, 1–8.

Brown, J. M., & Denney, H. I. (2007). Shifting attention into and out of objects: Evaluating the processes underlying the object advantage. *Perception & Psychophysics*, *69*(4), 606–618.

Brueggemann, A., & Gable, S. (2018). Preschoolers' selective sustained attention and numeracy skills and knowledge. *Journal of Experimental Child Psychology*, *171*, 138–147.

Casey, B., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological psychology*, *54*(1-3), 241–257.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2530–2546.

Cohen, L. B. (1969). Alternative measures of infant attention. In *Meeting of the Society for Research in Child Development*.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development*, *43*, 869–879.

Cohen, L. B., DeLoache, J. S., & Rissman, M. W. (1975). The effect of stimulus complexity on infant visual attention and habituation. *Child Development*, *46*, 611–617.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. In *Advances in child development and behavior* (Vol. 34, pp. 283–322). Elsevier.

Deng, W., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In *Principles of frontal lobe function* (pp. 466–503). Oxford University Press.

Diamond, A. (2006). The early development of executive functions. In *Lifespan cognition: Mechanisms of change* (Vol. 210, pp. 70–95). Oxford University Press.

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 11–43). American Psychological Association.

Diamond, A., Briand, L., Fossella, J., & Gehlbach, L. (2004). Genetic and neurochemical modulation of prefrontal cognitive functions in children. *American Journal of Psychiatry*, *161*(1), 125–132.

Doebel, S., Barker, J. E., Chevalier, N., Michaelson, L. E., Fisher, A. V., & Munakata, Y. (2017). Getting ready to use control: Advances in the measurement of young children's use of proactive control. *PLOS ONE*, *12*(4), e0175072.

Doebel, S., Dickerson, J. P., Hoover, J. D., & Munakata, Y. (2018). Using language to get ready: Familiar labels help children engage proactive control. *Journal of experimental child psychology*, *166*, 147–159.

Dubois, M., Aislinn, B., Moses-Payne, M. E., Habicht, J., Steinbeis, N., & Hauser, T. U. (2020). Exploration heuristics decrease during youth. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(5), 969–983.

Dye, M. W., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Research*, *50*(4), 452–459.

Egeland, J., & Kovalik-Gran, I. (2010). Measuring several aspects of attention in one test: The factor structure of Conners's continuous performance test. *Journal of Attention Disorders*, *13*(4), 339–346.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–177.

Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden*. Beltz.

Eng, C. M., Godwin, K. E., & Fisher, A. V. (2020). Keep it simple: Streamlining book illustrations improves attention and comprehension in beginning readers. *NPJ Science of Learning*, *5*(1), 1–10.

Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P., & Fisher, A. V. (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. *Journal of Experimental Child Psychology*, *138*, 126–134.

Fisher, A. V. (2019). Selective sustained attention: A developmental foundation for cognition. *Current Opinion in Psychology*, *29*, 248–253.

Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science*, *25*(7), 1362–1370.

Fisher, A. V., & Kloos, H. (2016). Development of selective sustained attention: The role of executive functions. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 215–237). American Psychological Association.

Fisher, A. V., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3-to 5-year-old children: Evidence from a new paradigm. *Journal of Experimental Child Psychology*, *114*(2), 275–294.

Franconeri, S., Jonathan, S., & Scimeca, J. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B*, *375*(1803), 20190502.

Harris, D. J., Vine, S. J., & Wilson, M. R. (2017). Is flow really effortless? The complex role of effortful attention. *Sport, Exercise, and Performance Psychology*, *6*(1), 103–114.

Holmboe, K., Fearon, R. P., Csibra, G., Tucker, L. A., & Johnson, M. H. (2008). Freeze-frame: A new infant inhibition task and its relation to frontal cortex tasks during infancy and early childhood. *Journal of Experimental Child Psychology*, *100*(2), 89–114.

Jensen, C. G., Vangkilde, S., Frokjaer, V., & Hasselbalch, S. G. (2012). Mindfulness training affects attention-or is it attentional effort? *Journal of Experimental Psychology: General*, *141*(1), 106–123.

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical psychology: Science and practice*, *10*(2), 144–156.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671.

Keebler, E., Kim, J., Thiessen, E. D., & Fisher, A. V. (2019). Does children's shape knowledge contribute to age-related improvements in selective sustained attention measured in a trackit task? In *CogSci* (pp. 1984–1990).

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S., & Kim, Y.-S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and Writing*, *27*(7), 1163–1188.

Kim, J., Singh, S., Thiessen, E. D., & Fisher, A. V. (2020). A hidden Markov model for analyzing eye-tracking of moving objects. *Behavior Research Methods*, *52*, 1–19.

Kim, J., Vande Velde, A., Thiessen, E. D., & Fisher, A. V. (2017). Variables involved in selective sustained attention development: Advances in measurement. In *Proceedings of the 39th annual conference of the cognitive science society*. Cognitive Science Society.

Kowler, E. (1995). Eye movements. *Visual Cognition*, *2*, 215–265.

Kwon, M.-K., Setoodehnia, M., Baek, J., Luck, S. J., & Oakes, L. M. (2016). The development of visual search in infancy: Attention to faces versus salience. *Developmental Psychology*, *52*(4), 537–555.

Lamy, D., & Egeth, H. (2002). Object-based selection: The role of attentional shifts. *Perception & Psychophysics*, *64*(1), 52–66.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900.

Laureiro-Martínez, D., Brusoni, S., & Zollo, M. (2010). The neuroscientific foundations of the exploration–exploitation dilemma. *Journal of Neuroscience, Psychology, and Economics*, *3*(2), 95–115.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, *12*(4), 163–169.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Transcription format and programs* (Vol. 1) Psychology Press.

Mahy, C. E., Mazachowsky, T. R., & Pagobo, J. R. (2018). Do verbal reminders improve preschoolers' prospective memory performance? It depends on age and individual differences. *Cognitive Development*, *47*, 158–167.

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, *79*(5), 1255–1274.

Monk, C. A., Boehm-Davis, D. A., Mason, G., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors*, *46*(4), 650–663.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Moyer, K. E., & Gilmer, B. v. H. (1955). Attention spans of children for experimentally designed toys. *The Journal of Genetic Psychology*, *87*(2), 187–201.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459.

Parmentier, F. B., & Andrés, P. (2010). The involuntary capture of attention by sound. *Experimental Psychology*, *57*, 68–76.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, *28*(6), 723–732.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Riccio, C. A., Reynolds, C. R., Lowe, P., & Moore, J. J. (2002). The continuous performance test: A window on the neural substrates for attention? *Archives of Clinical Neuropsychology*, *17*(3), 235–272.

Rosenshine, B. V., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Teacher Education*, *4*(1), 3–16.
Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350.
Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, *39*(5), 877–890.
Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: Themes and variations*. Oxford University Press.
Shallice, T., Stuss, D. T., Alexander, M. P., Picton, T. W., & Derkzen, D. (2008). The multiple dimensions of sustained attention. *Cortex*, *44*(7), 794–805.
Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487–518.
SMI (2009). *SMI: RED250 technical specification*. SensoMotoric Instruments.
Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4096–4108.
Spering, M., & Carrasco, M. (2015). Acting without seeing: Eye movements reveal visual processing without awareness. *Trends in Neurosciences*, *38*(4), 247–258.
Spering, M., Pomplun, M., & Carrasco, M. (2011). Tracking without perceiving: A dissociation between eye movements and motion perception. *Psychological Science*, *22*(2), 216–225.
Student, (1908). Probable error of a correlation coefficient. *Biometrika*, *6*, 302–310.
Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences*, *769*, 191–212.
Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area V4. *Neuron*, *61*(6), 952–963.
Tavassoli, A., & Ringach, D. L. (2010). When your eyes see more than you do. *Current Biology*, *20*(3), R93–R94.
Thomson, D. R., Besner, D., & Smilek, D. (2015). A resource-control account of sustained attention: Evidence from mind-wandering and vigilance paradigms. *Perspectives on Psychological Science*, *10*(1), 82–96.
Trick, L., Hollinsworth, H., & Brodeur, D. A. (2009). Multiple-object tracking across the lifespan: Do different factors contribute to diminished performance in different age groups. In *Computation, Cognition, and Pylyshyn* (pp. 79–99). MIT Press.
Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "catch the spies" task. *Cognitive Development*, *20*(3), 373–387.
Vales, C., Keebler, E., Kim, J., Girdhar, T., & Fisher, A. (2021). *Effects of labels and familiarity on sustained attention*. https://osf.io/kxbve/
Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science*, *18*(1), 65–79.
Vales, C., & Smith, L. B. (2018). When a word is worth more than a picture: Words lower the threshold for object identification in 3-year-old children. *Journal of Experimental Child Psychology*, *175*, 37–47.
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.
Wetzel, N., Scharf, F., & Widmann, A. (2019). Can't ignore-distraction by task-irrelevant sounds in early and middle childhood. *Child Development*, *90*(6), e819–e830.
Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision*, *10*(14), 29.

How to cite this article: Kim, J., Singh, S., Vales, C., Keebler, E., Fisher, A. V., & Thiessen, E. D. (2023). Staying and Returning dynamics of young children's attention. *Developmental Science*, *26*, e13410. https://doi.org/10.1111/desc.13410

APPENDIX A

A.1 | Re-evaluating staying after removing the first period of attention

Because the Target object is identified by a red circle prior the onset of each trial, participants tend to begin trial attending to the Target. As discussed in the main paper, if a participant's attention transitions sufficiently infrequently, the participant might exhibit above-chance Staying ability even in the absence of endogenous control over attention maintenance. To evaluate this possibility, this Appendix provides an additional post hoc analysis involving a modified version of our measure $\widehat{S}$ of Staying. Specifically, rather than computing average durations of every continuous period of attention maintenance on a single object, we excluded the first such period from the calculation (regardless of the object being attended during that first period). By construction, as with $\widehat{S}$, this alternative measure of Staying, which we denote $\widehat{S}'$ in the following sections, has a chance value of 0s under the null hypothesis that attention maintenance is not selective to the Target.

On the original dataset used in Experiment 1, the basic statistics of this alternative Staying measure across the 42 participants, analogous to the statistics provided for the original version of $\widehat{S}$ and other performance measures in Table 1 of the main paper, are as follows: the mean value was 0.88s (95% confidence interval (0.71s, 1.06s)) the standard deviation was 0.59s, and a two-sided *t*-test rejected the null hypothesis that the mean value was chance (0s) with $t(41) = 9.70, p < 0.001$. The mean $\widehat{S}'$ was not significantly lower than $\widehat{S}$ according to a paired two-sample *t*-test ($t(41) = 1.37, p = 0.179$).

On the dataset used in Experiment 2, the basic statistics of this alternative Staying measure across the 36 participants, analogous to the statistics provided for the original version of $\widehat{S}$ and other performance measures in Table 4 of the main paper, are as follows: the mean value was 0.98 s (95% confidence interval (0.79 s, 1.18 s)) the standard deviation was 0.64 s, and a two-sided *t*-test rejected the null hypothesis that the mean value was chance (0) s with $t(35) = 9.03, p < 0.001$. The mean $\widehat{S}'$ was not significantly lower than $\widehat{S}$ according to a paired two-sample *t*-test ($t(35) = 1.37, p = 0.279$).

Since, after the trial has started and participants have performed at least one object transition, there is no exogenous support to attending longer to the Target than to other objects, these analyses support the main paper's conclusion that participants' endogenously influence the durations of their attention on objects such that they attend to task-relevant (Target) objects for longer than to task-irrelevant (Distractor) objects.

A.2 | Validation of hmm with human coding

The main paper presented analyses based on an HMM algorithm developed by Kim et al. (2020) for inferring the object of a participant's

TABLE A1 Basic univariate statistics for distributions (across 42 participants) of each of the eye-tracking-based performance measures used in this study, computed from Human-Coded data.

Measure	Chance value	Mean (95% CI)	Std. Dev.	$t(41)$-value	p-value
$\hat{R}$	$1/6 = 0.1\overline{66}$	0.62 (0.55, 0.69)	0.23	12.59	<0.001***
$\hat{S}$	0	0.49s (0.30s, 0.68s)	0.62s	8.61	<0.001***
$\hat{S}'$	0	0.35s (0.27s, 0.43s)	0.26s	5.03	<0.001***

$^{***}p < 0.001$.

TABLE A2 Matrix of Pearson correlations between age and performance measures, computed from Human-Coded data.

	Age	Location Accuracy	Memory Accuracy	$\hat{R}$	$\hat{S}$
Age	–	0.532***	0.263	0.487***	0.273
Location Accuracy	–	–	0.285	0.547***	0.409**
Memory Accuracy	–	–	–	0.304	0.174
$\hat{R}$	–	–	–	–	0.450**
$\hat{S}$	–	–	–	–	–

Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).
$^{**}p < 0.005$, $^{***}p < 0.001$.

attention at 60 Hz. Kim et al. (2020) showed that their algorithm consistently agreed with judgements made by trained human coders, both in terms of the object, a participant was attending to a given frame and in terms of whether the participant transitioned between two objects on a given pair of consecutive frames.

While these validation analyses provide compelling justification for using the HMM algorithm to analyze eye-tracking data from the TrackIt experiment, the present paper is the first to use the HMM algorithm as a drawing conclusions of psychological interest. When using such an automated analysis tool, it is possible that the tool systematically differs from prior methods in a manner that might bias the conclusions of a study. In this section, to evaluate the possibility of such bias, we reran the analyses presented in the main paper using human-coded data instead of the HMM-coded data. As described below, the results successfully replicated the findings of the main paper, providing evidence for the utility of the HMM algorithm in a psychological study for the first time. Details of the human coding procedure can be found in Kim et al. (2020) under "Video coding procedure," and a script used by coders is available at https://osf.io/54kyd/.

Before giving the results of this analysis, we note a few differences between the human-coded dataset and the HMM-coded dataset used in the main paper. First, since the frame rate of the original dataset was quite high (60 frames/s), in order to reduce the onerous task of manually coding millions of frames over the dataset, the framerate was downsampled by a factor of 6 for human coders (i.e., coders labeled only every 6th frame, resulting in 10 labels per second). Second, in addition to the seven object states that the HMM assigned (nonmissing) frames, human coders were allowed to assign an additional "Off-Task" state for nonmissing frames during which the participant did not appear to be tracking any particular object. We assumed that transitions to and from the Off-Task state would correspond to attentional transitions, and therefore, for the purpose of computing $\hat{R}$ and $\hat{S}$, Off-Task was treated as an additional (7)th Distractor state.

A.2.1 | Comparison with chance values

As shown in Table A1, participants performed above chance according to each of our eye-tracking-based measures, suggesting that participants' SSA was endogenously supported through both Staying and Returning. These findings all replicate those in Tables 1 and 4 of the main paper.

A.2.2 | Correlations between measures

Compared to the results presented in the main paper, the only change in statistical significance was that Memory Accuracy was is no longer significantly correlated with either $\hat{R}$ or $\hat{S}$ (as shown in Table A2). Since we did not draw any conclusions regarding these relationships, this difference does not affect the conclusions of the main paper.

A.2.3 | Mediation analysis

As illustrated in Figure A1, the results of mediation analysis replicated those from the main analysis, with the indirect effect explaining an estimated 36% of common variation between Age and Location Accuracy. This mediation effect was still statistically significant ($p = 0.002$) according to the bootstrap procedure described in Preacher and Hayes (2008).

A.2.4 | Differential development of Returning and Staying over Age

As in Experiment 1, the correlation between $\hat{R}$ and Age was significantly greater than that between $\hat{S}$ and Age ($z = 1.904$, $p = 0.028$), suggesting that Returning develops more over this age range than does Staying.

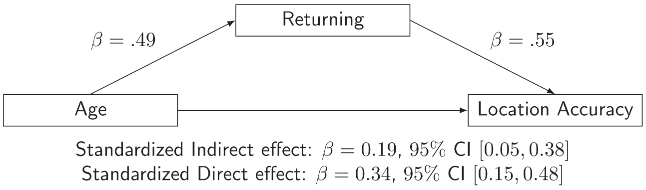

FIGURE A1 Standardized regression coefficients for the relationship between Age and Location Accuracy, as mediated by Returning, computed from Human-Coded data. Confidence intervals for indirect and direct effects were computed based on a bootstrap procedure with 10,000 bootstrap samples, as described by Preacher and Hayes (2008).

A.2.5 | Age-related changes in Average Tracking Duration

As in Experiment 1, participants' Average Tracking Duration increased significantly over Age ($\beta = 0.228$, $t(40) = 2.66$, $p = 0.011$).

A.2.6 | Effect of distance on Transitioning

The effect of distance on each type of transition probability, illustrated in Figure A2, was qualitatively similar to that in the main paper. Confidence bands were generally wider—this is expected, since, as described above, the total number of frames points is smaller (by a factor of 6) than in the HMM-coded dataset. Likely for the same reason, linear regression of transition frequencies over interobject distance, detailed in Table A3, showed similar overall patterns, but with lower significance levels. Specifically, the main finding, that the probability of transitions from Target to Distractors decreased significantly with interobject distance, was replicated. The probability of transitions from Distractors to the Target no longer changed significantly with interobject distance, although the regression coefficient was still positive. Again, the probability of transitions between Distractors did not significantly change with interobject distance.

To summarize, the analyses of the Human-Coded dataset reported in this appendix successfully replicated the key findings reported in the main paper. This supports the validity of the HMM approach proposed by Kim et al. (2020) for analyzing eye-tracking data in psychological studies.

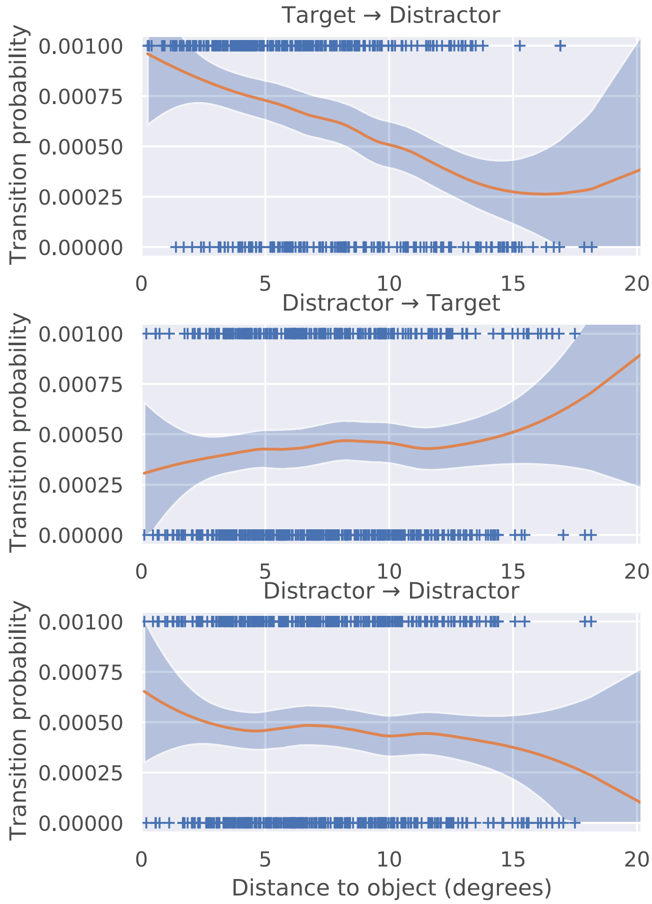

FIGURE A2 Nonparametric regression of empirical transition probability (i.e., probability of transitioning to a particular destination object given a particular source object) over distance between source and destination objects, for different source and destination object types, using the Human-Coded dataset. LOESS regression curves (orange) and 95% confidence bands (shaded blue) were computed by Python's SciPy package (Virtanen et al., 2020) using default parameters. Blue crosses show object distances corresponding to transitions (coded as 1) and nontransitions (coded as 0).

A.3 | Results on labeled condition of Experiment 2

Since the effect of shape labels in Experiment 2 was unrelated to any of the main questions of this paper, we presented only results from the Unlabeled condition of Experiment 2 in the main paper. This section presents results from the Labeled condition. All analyses were identical to those performed in Experiments 1 and 2 of the main paper. For brevity, we simply state the results; refer to Section 2.2 of Experiment 1 for motivation of each analysis. To summarize, the results were very similar to those presented in Experiments 1 and 2 of the main paper, with all of the main findings being replicated.

TABLE A3 Distributional statistics (across 42 participants in the Human-Coded dataset) of the correlation (across all frames within each participant) between interobject distance and each type of transition (with transitions coded as 1 and nontransitions coded as 0).

Transition type	Mean correlation	Standard deviation	$t(41)$-value	p-value
Target → Distractor	−0.16	0.25	−4.113	<0.001***
Distractor → Target	0.05	0.20	3.64	0.12
Distractor → Distractor	−0.01	0.23	−0.27	0.79

p-values are according to 1-sample t-tests for the null hypotheses that the mean correlations are 0.
***$p < 0.001$.

TABLE A4 Basic univariate statistics for distributions (across 42 participants) of each of the four performance measures in the Labeled condition of Experiment 2.

Measure	Chance value	Mean (95% CI)	Std. Dev.	t(41)-value	p-value
Location Accuracy	$1/16 = 0.0625$	0.47 (0.36, 0.58)	0.32	7.89	<0.001***
Memory Accuracy	$1/4 = 0.25$	0.87 (0.78, 0.97)	0.28	11.55	<0.001***
$\widehat{R}$	$1/4 = 0.25$	0.54 (0.46, 0.52)	0.24	8.82	<0.001***
$\widehat{S}$	0	0.89s (0.65s, 1.13s)	0.70s	7.28	<0.001***
$\widehat{S}'$	0	1.03s (0.66s, 1.40s)	1.10s	5.21	<0.001***

p-values are according to 1-sample t-tests for the null hypotheses that the means of the measures are equal to their chance values.
$^{***}p < 0.001$.

TABLE A5 Matrix of Pearson correlations between age and performance measures in the Labeled condition of Experiment 2.

	Age	Location Accuracy	Memory Accuracy	$\widehat{R}$	$\widehat{S}$
Age	–	0.621***	0.102	0.649***	0.334
Location Accuracy	–	–	0.726***	0.752***	0.570**
Memory Accuracy	–	–	–	0.683***	0.635**
$\widehat{R}$	–	–	–	–	0.477*
$\widehat{S}$	–	–	–	–	–

Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).
$^{*}p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.001$.

A.3.1 | Comparison with chance values

As shown in Table A4, participants performed above chance according to all four measures, suggesting that participants understood and were engaged in the task, and that SSA was endogenously supported through both Staying and Returning. These findings all replicate those in Table 1 of Experiment 1.

A.3.2 | Correlations between measures

Table A5 shows Pearson correlation coefficients between all four performance measures, as well as participant age, along with corresponding significance levels.

Location Accuracy improved strongly with Age, whereas Memory Accuracy was not significantly correlated with Age. Since Memory Accuracy is close to ceiling and much higher than Location Accuracy (see Table A4), we conclude that children are consistently able to encode and maintain the identity of the Target object, and that improvement in this ability fails to explain children's improving TrackIt performance over age. Both of our novel eye-tracking performance measures, $\widehat{R}$ and $\widehat{S}$, correlated positively with Location Accuracy, consistent with our expectation that Staying and Returning behaviors both support TrackIt performance. Additionally, $\widehat{R}$ increased significantly with Age, whereas $\widehat{S}$ did not. Finally, $\widehat{R}$ and $\widehat{S}$ were again significantly correlated.

These findings all replicate those of Table 2 in Experiment 1. The only notable difference (in terms of statistical significance) from Table 2 was that, here, Memory Accuracy was significantly correlated with Location Accuracy. Since, in this experiment, all objects had the same color, this may have been because encoding the distinct identity of the Target object was more difficult in this experiment and therefore played a bigger role in Location Accuracy performance than in Experiment 1; indeed, Memory Check performance ($M = 0.77$, $SD = 0.26$) was lower than in Experiment 1, ($M = 0.86$, $SD = 0.18$), although this difference was not statistically significant ($t(78) = 1.80$, $p = 0.076$).

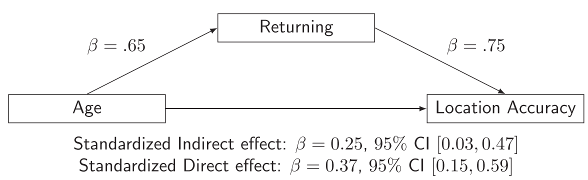

FIGURE A3 Standardized regression coefficients for the relationship between Age and Location Accuracy, as mediated by Returning, in Experiment 2. Confidence intervals for indirect and direct effects were computed based on a bootstrap procedure with 10,000 bootstrap samples, as described by Preacher and Hayes (2008).

A.3.3 | Mediation analysis

As illustrated in Figure A3, mediation analysis replicated results of Experiment 1, with the indirect effect explaining an estimated 51% (95% CI: (14%, 100%)) of common variation between Age and Location Accuracy. This mediation effect was statistically significant ($p = 0.003$) according to the bootstrap procedure described in Preacher and Hayes (2008).

When controlling for $\widehat{S}$, the results of the mediation analysis were similar: the estimated indirect effect was 0.25 (95% CI: (0.03, 0.47), the bootstrap p-value was $p = 0.034$, and the estimated proportion of mediation was 40% (95% CI: (9.66%, 100%)).

TABLE A6 Distributional statistics (across 42 participants) of the correlation (across all frames within each participant) between interobject distance and each type of transition probability (with transitions coded as 1 and non-transitions coded as 0), using data from the Labeled condition of Experiment 2.

Transition type	Mean correlation	Standard deviation	$t(41)$-value	p-value
Target → Distractor	−0.15	0.16	−6.32	<0.001***
Distractor → Target	0.058	0.15	2.46	0.007**
Distractor → Distractor	−0.040	0.15	−1.64	0.11

p-values are according to 1-sample t-tests for the null hypotheses that the mean correlations are 0.
$^{**}p < 0.005$, $^{***}p < 0.001$.

TABLE A7 Basic univariate statistics for distributions (across 42 participants) of each of the performance measures used in this study, computed over the subset of trials on which participants provided correct Memory Check responses.

Measure	Chance value	Mean (95% CI)	Std. Dev.	$t(41)$-value	p-value
Location Accuracy	$1/36 = 0.02\overline{77}$	0.42 (0.32, 0.51)	0.31	8.09	<0.001***
$\widehat{R}$	$1/6 = 0.1\overline{66}$	0.49 (0.43, 0.55)	0.20	10.16	<0.001***
$\widehat{S}$	0	0.92s (0.74s, 1.11s)	0.61s	9.76	<0.001***

$^{***}p < 0.001$.

A.3.4 | Differential development of Returning and Staying over Age

As in Experiment 1, the correlation between $\widehat{R}$ and Age was significantly greater than that between $\widehat{S}$ and Age ($z = 1.746$, $p = 0.040$), suggesting that Returning develops more over this age range than does Staying.

A.3.5 | Age-related changes in Average Tracking Duration

In contrast to Experiment 1 and the Unlabeled condition of Experiment 2, the increase in participants' Average Tracking Duration over Age was not statistically significant ($\beta = 1.51$, $t(40) = 1.122$, $p = 0.270$).

A.3.6 | Effect of distance on Transitioning

As shown in Table A6, the effect of distance on each type of transition probability was quite similar to that in the main paper.

A.4 | Results on subset of trials with correct memory response

The measures of Staying and Returning introduced in the main paper are designed around the assumption that the participant is tracking the correct Target object. However, it is possible that participants track another object for reasons other than failure of SSA (e.g., confusing the identity of the Target object). In this case, they may perform the cognitive behaviors of Staying and Returning towards that other object, but this would not be reflected in our Staying and Returning measures.

The ideal way to account for this possibility would be to include, in our analyses, only those trials on which the participant was attempting to sustain attention on the Target, rather than a Distractor. The most apparent proxy for this is the participant's Location Response. Unfortunately, since participants' mean Location Accuracy was 40%, analyzing only trials with correct Location Responses required us to discard most (60%) of our data, and the resulting correlation matrix (of Age, Returning, and Staying) contained no statistically significant entries. A power analysis found that the largest correlation in the matrix (which was that between Returning and Age) had power only about 18.2%. Thus, any interpretation of the null correlations over trials with correct Location Accuracy would be inconclusive.

We, therefore, utilized a less direct proxy, namely the accuracy of the participant's post-trial Memory Response, reasoning that correctly identifying the Target after the trial should correlate strongly with having tracked the Target during the trial. Since mean Memory Accuracy was quite high (86%), this resulted in discarding only a small proportion of trials. Below, we present the results of rerunning all of our main analysis of Staying and Returning on the subset of trials on which participants provided correct Memory Responses. To summarize, the results were very similar to those presented in Experiment 1 of the main paper, with all of the main findings being replicated.

A.4.1 | Comparison with chance values

As shown in Table A7, participants performed above chance according to each of our eye-tracking-based measures, suggesting that participants' SSA was endogenously supported through both Staying and Returning. These findings all replicate those in Tables 1 and 4 of the main paper.

A.4.2 | Correlations between measures

The matrix of correlations between performance measures, shown in Table A8, was nearly identical to that of Experiment 1 in the main paper.

A.4.2 | Mediation analysis

As illustrated in Figure A4, the results of mediation analysis replicated those from the main analysis, with the indirect effect explaining an estimated 42% of common variation between Age and Location Accuracy. This mediation effect was still statistically significant ($p = 0.001$) according to the bootstrap procedure described in Preacher and Hayes (2008).

TABLE A8 Matrix of Pearson correlations between age and performance measures, computed over the subset of trials on which participants provided correct Memory Check responses.

	Age	Location Accuracy	$\hat{R}$	$\hat{S}$
Age	–	0.539***	0.568***	0.334*
Location Accuracy	–	–	0.579***	0.485**
$\hat{R}$	–	–	–	0.443**
$\hat{S}$	–	–	–	–

Statistical significance was computed by a two-sided test of the null hypothesis that the correlation is 0 (Student, 1908).
*$p < 0.05$, **$p < 0.005$, ***$p < 0.001$.

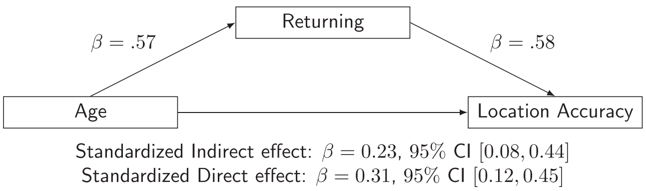

FIGURE A4 Standardized regression coefficients for the relationship between Age and Location Accuracy, as mediated by Returning, computed from Human-Coded data. Confidence intervals for indirect and direct effects were computed based on a bootstrap procedure with 10,000 bootstrap samples, as described by Preacher and Hayes (2008).

A.4.3 | Differential development of Returning and Staying over Age

As in Experiment 1, the correlation between $\hat{R}$ and Age was significantly greater than that between $\hat{S}$ and Age ($z = 2.007$, $p = 0.022$), suggesting that Returning develops more over this age range than does Staying.

A.4.4 | Effect of distance on Transitioning

As shown in Table A9, the effect of distance on each type of transition probability was quite similar to that in the main paper.

A.5 | Connections to the multiple object tracking (MOT) task

The TrackIt task shares many features with the Multiple Object Tracking (MOT) task, within which eye-tracking has been widely used to study visual attention (Meyerhoff et al., 2017). The MOT task differs from TrackIt primarily in two respects:

1. In MOT, the participant simultaneously tracks multiple (often 3–4) Target objects.
2. The Target and Distractor objects in MOT are visually indistinguishable. In particular, if the participant loses track of a Target object, they cannot recover the object based on its features (and hence the concept of Returning is difficult to establish in the context of MOT).

These factors make MOT much more challenging than TrackIt, and, while MOT has been used extensively with older children (Brockhoff et al., 2016; Dye & Bavelier, 2010; Trick et al., 2009), few studies have used it with children younger than 6 years (Trick et al., 2005). Additionally, in MOT, participants must track at least some of the Targets covertly, partly dissociating gaze and visual attention. These factors make it compelling to ask which findings in the MOT task extend to younger children performing the TrackIt task. This Appendix discusses a few connections between our findings and those from the MOT literature.

First, Franconeri et al. (2010) have argued that spacing between objects is the main determinant of MOT performance, and Bae and Flombaum (2012) have shown specifically that tracking errors in MOT occur primarily during (near-)collisions between objects. Our finding in the main paper that children tend to leave the Target when it (nearly) collides with other objects appears to mirror these findings in adults. Further work needs to be done to understand whether these phenomena share an underlying mechanism.

Second, Zelinsky and Todor (2010) have shown that transitions towards Target objects in MOT are often initiated in response to impending occlusion of the Target by a Distractor, an effect they refer to as "rescue saccades." Given the findings of Franconeri et al. (2010) and Bae and Flombaum (2012) noted above, these "rescue saccades" are likely a top-down response to a bottom-up signal that a Target is in imminent danger of being lost, and hence additional attentional resources should be allocated towards that Target. It is possible that Distractors threatening to occlude the Target could exogenously trigger "rescue saccades" to the Target in TrackIt. If some Distractor-to-Target (Returning) transitions are really "rescue saccades," this would contradict our hypothesis that Returning is initiated in a top-down manner. We, therefore, investigated the possibility of "rescue saccades" by testing if participants tracking a Distractor were more likely to return to the Target when the Target was very close to any other Distractor. We found that, when participants were tracking any Distractor, on average over participants, the distance from the Target to its

TABLE A9 Distributional statistics (across 42 participants) of the correlation (across all frames within each participant) between interobject distance and each type of transition probability (with transitions coded as 1 and non-transitions coded as 0), using only the subset of trials on which the participant provided a correct Memory Response. *p*-values are according to 1-sample *t*-tests for the null hypotheses that the mean correlations are 0.

Transition type	Mean correlation	Standard deviation	$t(41)$-value	*p*-value
Target → Distractor	−0.12	0.13	−5.78	<0.001***
Distractor → Target	0.05	0.13	2.46	0.018*
Distractor → Distractor	−0.008	0.11	−0.52	0.61

*$p < 0.05$, ***$p < 0.001$.

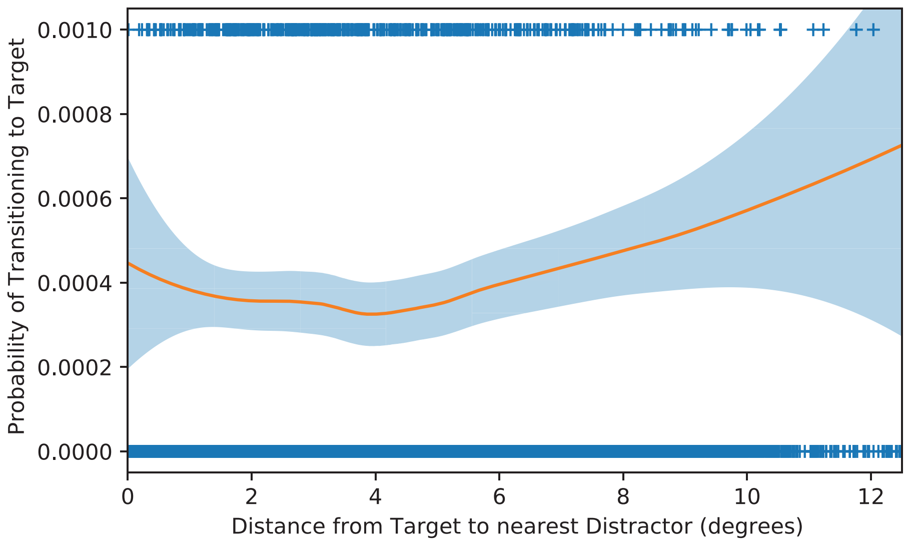

FIGURE A5 Nonparametric regression of empirical Distractor-to-Target transition probability (i.e., probability of transitioning to the Target on the next frame when currently tracking a Distractor) over distance between the Target and its nearest Distractor object. LOESS regression curves (orange) and 95% confidence bands (shaded blue) were computed by Python's SciPy package Virtanen et al. (2020) using default parameters. Blue crosses show Target-to-nearest-Distractor distances corresponding to transitions (coded as 0.001) and nontransitions (coded as 0).

nearest Distractor had no significant effect on the participant's probability of returning to the Target on the next frame ($t(41) = 1.69$, $p = 0.098$). Figure A5 shows a nonparametric regression of the Distractor-to-Target probability over the distance from the Target to its nearest Distractor. There is a slight increase in probability near 0, but this effect is very small, and even if we consider only frames during which the Target was fairly close to its nearest Distractor (<4°), the effect is not statistically significant ($t(41) = -1.25$, $p = 0.219$).

To conclude, we did not find evidence of "rescue saccades." This could be caused by one of several possible factors. A first possibility is that, while tracking a Distractor, participants did not track the Target covertly, which would be necessary in order to initiate a "rescue saccade." Another possibility is that participants did sometimes track the Target covertly, but did not deem "rescue saccades" necessary, because, in contrast to MOT, the Target in TrackIt is visually distinguishable from each Distractor and can thus always be recovered later, even if the participant loses track of it. A final possibility is simply that the cognitive mechanism underlying "rescue saccades" has not yet developed in children aged 3.5–6 years.

A.6 | Effect of interobject distance on transition probabilities over age

In the main paper, we reported finding the probability of transitioning from the Target to a Distractor decreased with the distance between the Target and that Distractor ("interobject distance"). Additionally, we hypothesized that this effect was caused by a decrease in the bottom-up salience of Distractors further away from the Target and center of the visual field. Under this hypothesis, since top-down inhibition is believed to play an increasing role in moderating the effect of bottom-up salience on attention over the course of development (Diamond, 2006; Ruff & Rothbart, 2001), then we might expect this effect to decrease in magnitude with age. We, therefore, regressed the correlations between interobject distance and each type of transition probability (computed as described in Figure 4 of the main paper) over participant Age. In Experiment 1, as shown in Table A10, this yielded no statistically significant effect for any transition type, although the coefficients for both Target-to-Distractor and Distractor-to-Distractor were trending positive.

In Experiment 2, on the other, as shown in Table A11, this effect was significant for Target → Distractor, but still not for other transition types. Overall, although consistent with a mitigating effect of Age on the change in transition probabilities over interobject

TABLE A10 Linear regression (across participants) of correlation (across all frames within each participant) between interobject distance and each type of transition probability, over Age.

Transition type	Regression coefficient β	$t(40)$	*p*-value
Target → Distractor	0.132	1.599	0.118
Distractor → Target	0.012	0.244	0.809
Distractor → Distractor	0.120	1.724	0.093

TABLE A11 Linear regression (across participants) of correlation (across all frames within each participant) between interobject distance and each type of transition probability, over Age, using data from the Unlabeled condition of Experiment 2.

Transition type	Regression coefficient β	$t(34)$	*p*-value
Target → Distractor	0.546	2.376	0.022*
Distractor → Target	0.180	0.743	0.462
Distractor → Distractor	0.240	0.977	0.335

*$p < 0.05$.

distance, we consider these results inconclusive due to limited power (e.g., about 30.3% for Target → Distractor at the observed effect size in Experiment 1).

A.7 | Additional exogenous influences on attentional transitioning

In the main paper, we investigated the effect of one exogenous factor, inter-object distance, that we proposed might influence transitions, especially transitions towards Distractor objects. However, many other exogenous factors could potentially influence attentional transitions.

One possible exogenous influence is the shape of an object, which could influence transitions in several possible ways. For example, more complex objects (e.g., Crescent or Arrow) might be more salient, or more novel, than simpler or more familiar objects (e.g., Circle or Square), possibly promoting transitions towards them. To investigate this possibility, as well as the possibility of differential effects for transitions to the Target (Returning) versus transitions to Distractors, we ran a (post hoc) two-way ANOVA, including Object Shape (a categorical variable, having nine levels in Experiment 1 and six levels in Experiment 2) and Object Type (Target or Distractor) as factors.

As shown in Table A12, on both datasets used in Experiments 1 and 2, this yielded significant main effects of Object Type (transitions to the Target were much more likely than transitions to Distractors), but no significant main effects of object shape or interaction between Object Shape and Object Type.

We note that the TrackIt objects were not designed to differ significantly in terms of bottom-up salience; indeed, in most studies using TrackIt, this would simply add an undesired source of trial-to-trial variability. Hence, it is not surprising that we did not observe any significant effect of object shape, and future studies should use stronger manipulations to investigate effects of bottom-up salience on difference types of attentional transitions.

TABLE A12 Results of two-way ANOVA investigating the effects of Object Shape and Object Type on probability of transitioning to an object.

Experiment	Factor	Degrees of freedom	F	p
Experiment 1	Object shape	8	1.82	0.070
	Object type	1	222.99	<0.001***
	Object shape × Type	8	1.77	0.079
Experiment 2	Object shape	5	0.48	0.79
	Object type	1	136.47	<0.001***
	Object shape × Type	5	0.44	0.82

***$p < 0.001$.