


The Development of Attention to Objects and Scenes: From Object-Biased to Unbiased

Kevin P. Darby *Department of Psychology, University of Virginia*

Sophia W. Deng

Department of Psychology, University of Macau

Dirk B. Walther

*Department of Psychology, University of Toronto and Samsung Artificial Intelligence Center Toronto*Vladimir M. Sloutsky *Department of Psychology, Ohio State University*

Selective attention is the ability to focus on goal-relevant information while filtering out irrelevant information. This work examined the development of selective attention to natural scenes and objects with a rapid serial visual presentation paradigm. Children ($N = 69$, ages 4–6 years) and adults ($N = 80$) were asked to attend to either objects or scenes, while ignoring the other type of stimulus. A multinomial processing tree model was used to decompose selective attention into focusing and filtering components. The results suggest that attention is object-biased in children, due to difficulty filtering attention to goal-irrelevant objects, whereas attention in adults is relatively unbiased. The findings suggest important developmental asymmetries in selective attention to scenes and objects.

Our visual world is made up of scenes and objects, and we can attend selectively by focusing on specific objects or aspects of scenes that are most relevant to our current goals, and by filtering or inhibiting aspects that are less relevant (Broadbent, 1958; Lachter, Forster, & Ruthruff, 2004). In this work, we examine the development of selective visual attention to natural scenes and objects under potential interference from goal-irrelevant stimuli.

The Development of Attention to Scenes and Objects

A visual scene often contains many objects and more global scene properties. For example, a city scene is likely to contain people, signs, and

buildings, coupled with more spatially distributed properties, such as the presence of primarily vertical contours. However, only a subset of information may be processed at any given time by *focusing* on goal-relevant information and *filtering*, or inhibiting, goal-irrelevant information. Attention may be selectively focused on individual objects or distributed more broadly to process multiple spatially distributed objects or properties of the overall scene (Treisman, 2006). Because attention is a limited resource, distributing attention to process the scene as a whole may result in reduced processing of individual objects, and vice versa.

Although selective attention is used to focus on or prioritize goal-relevant information, sometimes goal-irrelevant information is not completely filtered out, causing interference. For example, previous work with adults has shown that recognizing an object is more difficult when the object is surrounded by a scene, particularly when the scene is semantically incongruent with the object (e.g., a football player in a cathedral; Davenport & Potter, 2004). Similarly, the presence of a salient object can interfere with the recognition of natural scenes (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007).

This work was supported by National Institutes of Health Grants R01HD078545 and P01HD080679 to Vladimir M. Sloutsky, and an NSERC Discovery Grant (#498390) and the Canadian Foundation for Innovation (#32896) to Dirk B. Walther.

Conflict of interest: Dirk B. Walther is a Visiting Professor at the Samsung Artificial Intelligence Center Toronto. The work in this manuscript is separate from his work for Samsung. It was performed exclusively in his role as a faculty member at the Ohio State University and the University of Toronto. The other authors declare no conflicts of interest.

Data Availability Statement: The data that support the findings of this study are available at <https://osf.io/5pva9/>.

Correspondence concerning this article should be addressed to Vladimir M. Sloutsky, Department of Psychology, The Ohio State University, 247 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210. Electronic mail may be sent to sloutsky.1@osu.edu.

Interference from irrelevant scenes or objects is likely to vary across development due to changes in selective attention. Past work has suggested that selective attention develops substantially during childhood, particularly between 4 and 7 years of age (see Hanania & Smith, 2010; Plude, Enns, & Brodeur, 1994, for reviews). Young children are known to have difficulty selectively attending to stimuli, showing more distributed attention compared to adults in a variety of tasks, including visual search (Pastò & Burack, 1997; Plebanek & Sloutsky, 2017) and category learning (Deng & Sloutsky, 2015, 2016). For example, Plebanek and Sloutsky (2017) presented 4- to 5-year-old children and adults with a change-detection task for stimuli including task-relevant and irrelevant features. Whereas adults were better able to detect changes in the task-relevant feature, children performed better than adults when detecting changes in the task-irrelevant features, suggesting that children's attention was less selective. In a follow-up study, Sloutsky and Plebanek (under review) manipulated demands on the focusing and filtering components of selective attention, and found evidence that developmental changes in selective attention may be primarily driven by changes in filtering.

There is also evidence from studies examining the development of selective attention for objects with a rapid serial visual presentation (RSVP) paradigm (Day & Stone, 1980). In this study, 5-year-olds, 8-year-olds, and adults were presented with an image of a target object, followed by a test object that was either the same or different from the target. The test object was either presented immediately after the target, or following a stream of rapidly presented distractor objects. The ability to detect a repeated test object despite potential interference from the RSVP stream improved across development, suggesting a greater ability to filter out irrelevant objects. It remains unknown from this work, however, how selective attention for scene information develops, or how it may compare to selective attention for object information.

The presence of attentional interference when attempting to selectively attend to scene or object information may provide insight into how these stimuli are processed across development. For example, a tendency to distribute attention to the scene as a whole may make it more difficult to attend to specific objects. In general, natural scenes are likely to contain more spatially distributed information than individual objects. Therefore, given that young children often exhibit more distributed visual attention (Deng & Sloutsky, 2015,

2016; Pastò & Burack, 1997; Plebanek & Sloutsky, 2017, 2019), one hypothesis is that children may be likely to attend to global scene information even when attempting to focus on specific objects, whereas objects may interfere less (or not at all) with scene processing.

Another possibility is that children's distributed attention is offset by an attentional bias toward objects. This possibility is supported by work suggesting that objects are units of attention in infants and children (Bulf & Valenza, 2013; Spelke, 1990; Werchan, Lynn, Kirkham, & Amso, 2019), as well as in adults (Driver & Baylis, 1989; Duncan, 1984; Egly, Driver, & Rafal, 1994; Spelke, 1990; Treisman, Kahneman, & Burkell, 1983; see Chen, 2012, for a review). For example, Egly et al. (1994) cued participants to the location of a target on one of two rectangles, and found that adult participants were better able to find an invalidly cued target when it was on the same rectangle as the cue, compared to when it was the same distance from the cue but on a different rectangle. This suggests that attention may be deployed on the basis of objects, such that features within objects are more readily processed than features distributed across objects. If attention is biased toward objects in children, objects may attract attention regardless of goals, such that goal-irrelevant objects may interfere with the processing of scenes, whereas scenes may interfere little with object processing.

These potential biases toward scenes or objects could be driven by focusing, filtering, or a combination of these processes. For example, interference from objects could be due to a failure to filter irrelevant object information, or a failure to focus on goal-relevant scenes. In the reported research, we examine these issues using a combination of experimentation and modeling.

Current Work

In this work, we investigated how the ability to attend selectively to natural scenes and objects develops. Specifically, we (a) asked how preschool-aged children and adults extract and integrate scene and object information and (b) examined the attentional mechanisms of focusing on goal-relevant information and filtering out goal-irrelevant information. To do so, we used an RSVP paradigm, in which participants were shown a target stimulus of an object superimposed on a scene, and were then shown a rapid stream of other object-scene stimuli that either contained the same stimulus as the target or not. Attention was manipulated between

subjects: participants were instructed to pay attention to either the object or the scene of the target stimulus, and respond to whether the target object or scene was repeated in the RSVP stream, while ignoring the other type of stimulus.

To gain insight into the contributions of attentional focusing on task-relevant and filtering of task-irrelevant information, we developed a multinomial processing tree (MPT) model. MPT models are used to estimate how latent cognitive processes contribute to the frequency of different categorical responses based on probability (Batchelder & Riefer, 1999). Importantly, MPT models do not attempt to quantify on a cognitive or biological level *how* these processes are implemented in the brain, but assume that mechanisms contribute to performance in ways hypothesized in the model. This is in contrast to standard statistical models, such as analysis of variance (ANOVA) or regression models, which are atheoretical and do not address mechanisms contributing to performance without additional inferences on the part of the researcher (Lewandowsky & Farrell, 2011). MPT models have been productively used in prior developmental work to estimate the contribution of cognitive processes such as memory binding and semantic clustering to task performance (Horn, Bayen, & Michalkiewicz, 2020; Yim, Dennis, & Sloutsky, 2013). The MPT model we developed for the current work parameterized attention to relevant as well as irrelevant streams of information, allowing us to gain insight into developmental differences in focusing and filtering of scene and object information.

The question of interest was how well children and adults would be able to focus on goal-relevant scene or object information and filter out irrelevant information. Given prior evidence of more selective attention in adults compared to preschool-aged children (Deng & Sloutsky, 2015, 2016; Pastò & Burack, 1997; Plebanek & Sloutsky, 2017, 2019), we expected to find confirmatory evidence of developmental improvements in filtering (and potentially focusing) aspects of selective attention, given work suggesting that filtering may be an especially important contributor to the development of selective attention (Sloutsky & Plebanek, under review). The prior literature provided a less clear prediction for the difference between scene and object processing. As discussed above, one exploratory hypothesis is that children would have difficulty inhibiting attention to irrelevant scenes, as scene information is more spatially distributed, and a great deal of work has suggested distributed attention in children (Deng & Sloutsky, 2016; Plebanek & Sloutsky,

2017, 2019). An alternative exploratory hypothesis is that objects are units of attention, in which case objects may be expected to produce greater attentional interference than scenes.

Method

Participants

Ninety-four preschool-aged children ($M_{\text{age}} = 5.3$ years, $SD = .3$, range = 4.6–6.3; 39 females, 55 males) were recruited from preschools and day cares in Columbus, Ohio. Children were tested in a quiet room in their preschool or day care and received stickers for participating. Eighty-two adults (53 self-identified as female, 28 as male, and 1 as other) participated in exchange for partial course credit and were tested in a quiet laboratory room. Data collection followed all ethical guidelines set forth by The Ohio State University's Human Research Protection Program. All child participants verbally assented to participate; written consent was provided by a parent of all children and by all adults. Data were collected between June 2016 and February 2017.

Materials

Stimuli consisted of photographs of real-world scenes and objects, with five categories of both scenes (beaches, streets, offices, mountains, and kitchens) and objects (trees, cats, cars, slides, and people). Each category contained (a) 10 exemplars that were used in the main experiment and (b) an additional three exemplars that were used in a short training block at the beginning of the session.

The scene images were each 800×600 pixels. The object images were presented on transparent backgrounds, and each object contained approximately 5,000 nontransparent pixels. The experimental stimuli were formed by superimposing an object image on a scene image. The spatial location of the object on the scene was randomized using a 16×12 grid that was invisible to the participant. Stimuli were followed by a perceptual mask for 200 ms. Masks were generated by synthesizing random, scene-like textures separately for each color channel (red, green, and blue) using the texture synthesis method of Portilla and Simoncelli (2000). To create each mask, we averaged six random textures, each synthesized from a randomly chosen photograph from a separate set of scene images. We precomputed five such masks and presented each mask once on every trial.

Procedure and Design

At the beginning of the experiment, participants were instructed that they would see images of scenes and objects, and that they would need to pay attention to only scenes *or* objects, while ignoring the other type of stimulus throughout the entire experiment, in a between-subject manipulation of attention. Participants were first trained to match two images with the same relevant stimulus (i.e., a scene or object, depending on condition), ignoring the other, irrelevant stimulus. Participants were then provided with extensive instructions on how to perform the RSVP task, as well as a short training block in which they received response feedback (see Appendix A for details).

On each experimental trial, participants were shown a target scene and object, first separately (one at a time in a randomized order), and then together, with the object superimposed on the scene. The timing of these target stimulus presentations was self-paced by adults and controlled by the experimenter for children. Following the presentation of the target image, an RSVP stream of four images was flashed on the screen: one test image and three distractor images. The order of the distractor and test images was randomized for each trial. The scenes and objects in the distractor images were randomly selected from the stimulus set, except that distractor stimuli could not come from the same category as the scene or object presented in the target image.

Each image in the RSVP stream was presented for 800 ms to children and for 200 ms to adults. Pilot testing suggested that this difference in the stimulus duration was necessary to avoid ceiling and floor effects on performance. For both age groups, mask images were presented for 200 ms at the onset of the RSVP stream, as well as following each experimental stimulus in the stream, in an order randomized for each trial (see Figure 1).

Participants were instructed to watch the RSVP stream for a repetition of the relevant stimulus (object or scene) from the target image: if the relevant target stimulus was repeated, participants were asked to respond “old” by clicking either the left or right button on the mouse as quickly as possible (the buttons were randomly assigned for each participant). If the target stimulus was not repeated, participants were instructed to wait until the end of the stream (at which point a large question mark was presented on the screen), and then click the other button on the mouse to respond “new.” As is often done in RSVP paradigms (Gerson, Parra, & Sajda, 2005; Touryan, Gibson, Horne, & Weber,

2011; Vierck & Miller, 2006), we asked participants to make “old” responses as soon as possible. This was done to minimize potential memory interference that could arise from subsequently presented stimuli, and because we felt it would be more natural for participants to respond immediately. However, it was necessary for participants to wait until the end of the stream to make a “new” response so that they could observe all of the stimuli before determining that the target stimulus was absent. The RSVP stream was terminated immediately upon a participant’s response. The stream ended upon response because we reasoned that children may have difficulty sitting through additional images following their response. If a response was not provided during the RSVP stream, the question mark appeared until a response was made.

Test images were *congruent* when both test object and scene were old (i.e., matched the target) or both were new, so that attention to either or both of the stimulus types would result in the same response. They were *incongruent* when one of the stimuli was old and the other new, thus requiring different responses depending on the attentional focus condition.

The experiment therefore had a 2 (Age: Children vs. Adults) \times 2 (Attentional focus: Attention-to-scenes vs. Attention-to-objects) \times 2 (Congruency: Congruent vs. Incongruent) design, with Age and Attentional focus as between-subject factors and Congruency as a within-subject factor. Participants were presented with three blocks of 20 trials, for a total of 60 trials. Each of the four trial types—congruent old, congruent new, incongruent old, incongruent new—included 15 trials across the experiment (5 in each block). The experiment lasted approximately 40 min for children and 20 min for adults.

Results and Discussion

We focused our analyses on participants who (a) completed the task, and (b) clearly understood and followed task instructions. To this end, we excluded 15 children who did not complete the task, and an additional 10 children who failed to perform at above-chance accuracy on congruent trials, as determined by a one-tailed binomial test. We also excluded two adults due to failing to meet the accuracy criterion. The final sample, then, included 69 children ($M_{\text{age}} = 5.3$ years, $SD = .3$, range 4.8–6.3; 39 females, 30 males), and 80 adults (51 self-identified as female, 28 as male, and 1 as other). Of these participants, 35 children and 46 adults were

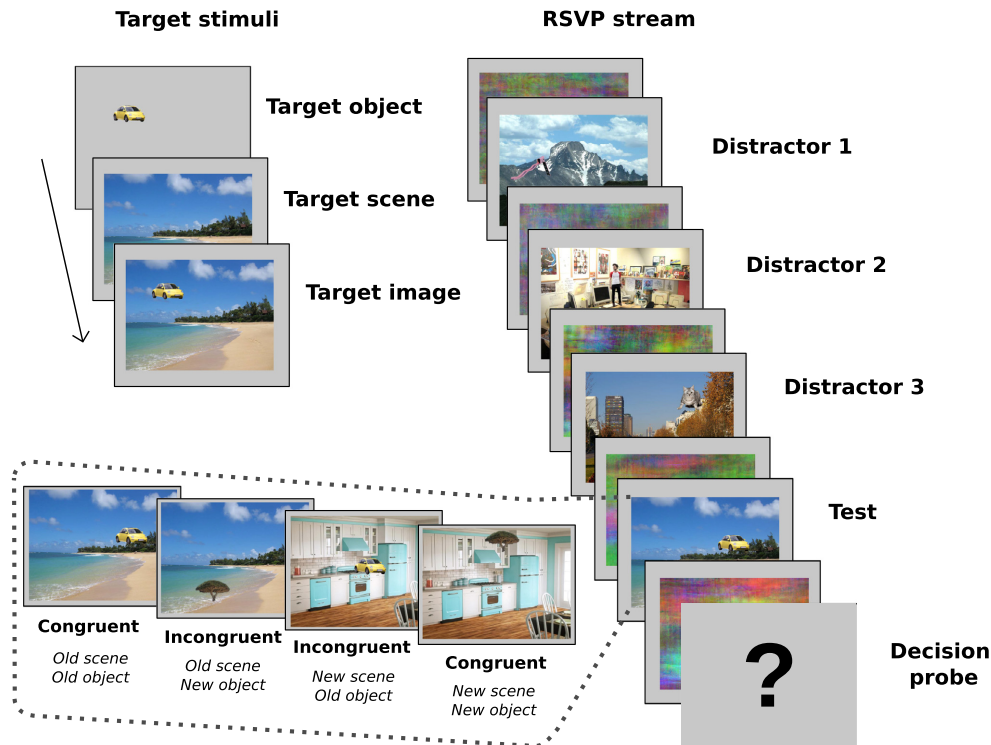


Figure 1. Experimental procedure. On each trial, participants were shown a target object and scene, and then an RSVP stream of images containing a test image and 3 distractor images, in an order randomized for each trial. The test image included a scene and object that were each either the same or different as presented in the target image (inset). The spatial location of the object was randomized for every image in the stream, including the test image. Participants were instructed to respond “old” immediately if they saw a repetition of the target object or scene in the RSVP stream, or to respond “new” after the stream had completed. The RSVP stream was terminated immediately upon a response; if no response was made before the end of the stream, a question mark appeared until a response was made. [Color figure can be viewed at wileyonlinelibrary.com]

randomly assigned to the attention-to-scenes condition and 34 children and 34 adults were randomly assigned to the attention-to-objects condition.

To investigate selective attention and potential attentional biases to scenes and objects, we compared performance on congruent and incongruent trials in the attention-to-objects and attention-to-scenes conditions. We reasoned that an attentional bias toward scene or object information would result in attending to that kind of information even when it was irrelevant. For example, an attentional bias toward objects could manifest as difficulty inhibiting attention to objects when they are task-irrelevant, such that performance would be lower on incongruent trials in the attention-to-scenes condition than in the attention-to-objects condition.

We measured performance by calculating d' —a signal detection statistic that provides bias-free estimates of discrimination of old (i.e., repeated) stimuli from new (i.e., not repeated) stimuli (Stanislaw & Todorov, 1999). The metric d' was calculated as $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$, where Z

signifies the inverse of the Gaussian cumulative distribution function. The observed d' performance for children and adults in both stimulus conditions are presented in the bar plots in Figure 2A.

Observed data presented in Figure 2A were submitted to a 2 (Age: Child vs. Adult) \times 2 (Attentional focus: Attention-to-objects vs. Attention-to-scenes) \times 2 (Congruency: Congruent vs. Incongruent) mixed ANOVA with Age and Attentional focus as between-subject factors and Congruency as a within-subject factor. The full results of this ANOVA are presented in Table 1. Importantly, there was a significant three-way interaction among all factors, $p = .004$, suggesting an age difference in interference effects when attempting to attend to scenes versus objects.

To better understand this three-way interaction, we then performed separate 2 (Attentional focus: Attention-to-scenes vs. Attention-to-objects) \times 2 (Congruency: Congruent vs. Incongruent) analyses of variance (ANOVAs) on performance in children and adults (see Table 1). In children, there was a significant

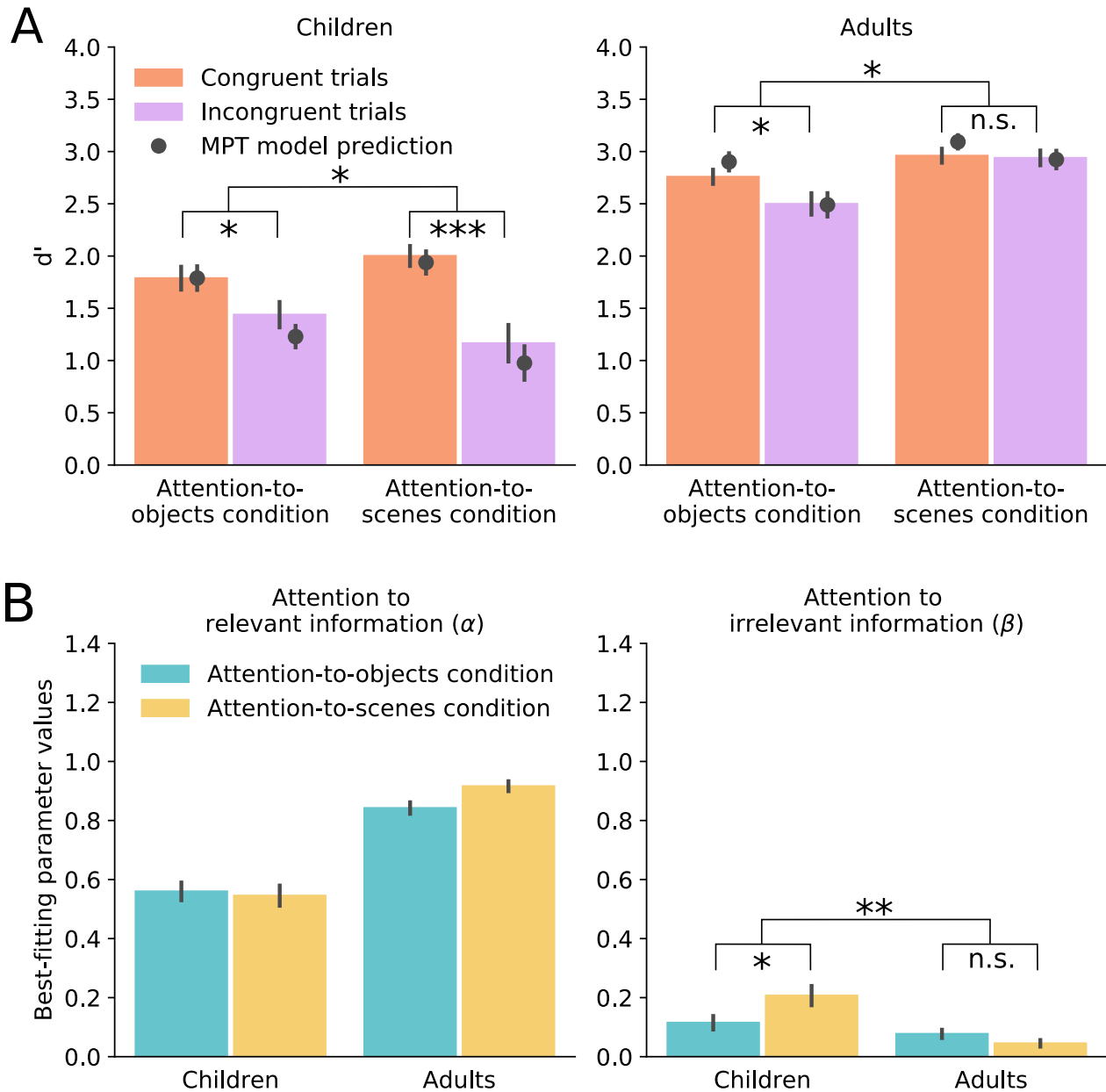


Figure 2. Observed and MPT model-predicted performance and MPT model parameter results. Performance was measured by response discrimination between old and new stimulus detection for each condition and age group (A). Model-predicted performance was calculated by generating performance in every condition using the best-fitting parameter values for each participant. To gain insight into focusing and filtering, we analyzed best-fitting parameter values of an MPT model estimating contributions of attention to task-relevant and irrelevant information (B). Error bars represent standard errors of the mean. * $p < .05$, ** $p < .01$, *** $p < .001$, n.s. $p > .05$. [Color figure can be viewed at wileyonlinelibrary.com]

interaction between these factors, $p = .04$, indicating a greater interference effect from irrelevant objects than from irrelevant scenes. The difference in performance between congruent and incongruent trials was strong in the attention-to-scenes condition, post hoc paired-samples t test, $t(34) = 4.67$, $p < .001$, $d = .79$, whereas it was weak in the attention-to-objects condition,

$t(33) = 2.42$, $p = .02$, $d = .42$. The interaction was also significant in adults, $p = .04$. However, in contrast to children, the difference between congruent and incongruent trials was not significant in the attention-to-scenes condition, $t(45) = 0.31$, $p = .76$, $d = .05$, but was significant in the attention-to-objects condition, $t(33) = 2.75$, $p = .01$, $d = .47$, indicating some

Table 1
ANOVA Results on Observed Performance

ANOVA	Factor	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
Age (Child vs. Adult) × Congruency	Age	(1, 145)	203.41	< .001	.58
(Congruent vs. Incongruent) × Attentional Focus (Attention-to-scenes vs. Attention-to-objects)	Congruency	(1, 145)	35.20	< .001	.20
	Attentional focus	(1, 145)	3.04	.09	.009
	Age × Congruency	(1, 145)	13.41	< .001	.09
	Age × Attentional Focus	(1, 145)	4.43	.04	.03
	Congruency × Attentional Focus	(1, 145)	1.01	.32	.007
	Age × Congruency × Attentional Focus	(1, 145)	8.62	.004	.06
<i>Children</i>	Congruency	(1, 67)	26.39	< .001	.28
Congruency (Congruent vs. Incongruent) × Attentional Focus (Attention-to-scenes vs. Attention-to-objects)	Attentional focus	(1, 67)	0.05	.83	.00
	Congruency × Attentional Focus	(1, 67)	4.45	.04	.06
<i>Adults</i>	Congruency	(1, 78)	6.12	.02	.07
Congruency (Congruent vs. Incongruent) × Attentional Focus (Attention-to-scenes vs. Attention-to-objects)	Attentional focus	(1, 78)	11.60	.001	.13
	Congruency × Attentional Focus	(1, 78)	4.44	.04	.05

Note. Statistically significant results are displayed in bold font. ANOVA = analysis of variance.

interference from scenes. However, because this interference was weak, with a small effect size comparable to that of the interference effect in the same condition in children, we caution that this effect does not provide strong evidence of a bias toward scenes in adults. We therefore conclude that adults are likely relatively unbiased in their selective attention between objects and scenes. The behavioral results shown in Figure 2A therefore suggest a developmental asymmetry, whereby children experienced strong attentional interference from objects, and adults showed some interference from scenes. Importantly, whereas interference from scenes was comparable in children and adults, interference from objects was substantially stronger in children.

One possible explanation for greater attentional interference from objects than scenes in children is that children were more familiar with object categories than scene categories. If this were the case, we would expect children to perform better on this task when attending to objects in the absence of scenes than vice versa. To investigate this, we performed a separate baseline experiment, in which only the task-relevant stimulus type was presented, with children and adults. There was no effect of stimulus type and no interaction with age (see Appendix B), which does not support the familiarity account.

We suggest instead that the behavioral results of the main experiment likely indicate developmental differences in attention. The finding that children showed greater interference from objects than

scenes suggests an attentional bias toward objects. In contrast, adults showed weak interference stemming from scenes that was comparable to that in children. There are different possibilities for the source of these differences. Specifically, it is possible that differences arise from the ability to filter irrelevant object information. Indeed, previous research suggests that, in general, filtering undergoes a great deal of developmental change between the ages of five years and adulthood (Wendelken, Baym, Gazzaley, & Bunge, 2011), and may be the primary source of developmental change in selective attention (Sloutsky & Plebanek, under review). However, it is also possible that there are important differences in focusing attention on *relevant* scene information. We next estimate these attentional processes by applying an MPT model of performance.

MPT Model of Performance

To estimate contributions of focusing and filtering on children's and adults' performance, we used an MPT model. For this model, we hypothesized that two latent processes gave rise to performance: attention to task-relevant stimuli (due to focusing, estimated by parameter α), and attention to task-irrelevant stimuli (due to the absence of filtering, estimated by parameter β). The model assumes that different attentional patterns will result in different responses. For congruent trials, we assumed that attention to either or both types of stimuli would

result in the correct response, whereas attention to neither type of stimulus would result in guessing. On incongruent trials, attention only to the task-relevant stimulus would be expected to result in the correct response, whereas attention only to the irrelevant stimulus would be expected to result in the incorrect response. If attention was paid to both types of stimuli, we assumed that participants would equally guess between the correct and incorrect responses, since both responses were supported by the target image. Likewise, when attention was paid to neither type of stimulus, we expected participants to guess. The α and β model parameters estimate the extent of attention to task-relevant and -irrelevant stimuli, and range from 0, indicating no attention, to 1, indicating full attention.

We estimated the values of the α and β parameters using a Bayesian approach. We fitted the model to all of the data of each individual participant independently, and estimated the maximum a posteriori, or the best-fitting value, of each parameter, for each participant (see Figure 2B). See Appendix C for a full description of the model and model-fitting procedures.

To assess whether there were differences in the contribution of attention to relevant and irrelevant stimuli between age groups and stimulus conditions, we performed a 2 (Age: Child vs. Adult) \times 2 (Attentional focus: Attention-to-objects vs. Attention-to-scenes) \times 2 (Parameter: α vs. β) mixed ANOVA with Age and Attentional focus as

between-subject factors and Parameter as a within-subject factor (see Table 2 for full results). Importantly, the three-way interaction was significant, $p = .002$, suggesting that the contribution of attention to relevant and irrelevant stimuli (estimated by the parameters) differed between age groups and attention conditions.

To better understand this interaction, we performed separate 2 (Age: Child vs. Adult) \times 2 (Attentional focus: Attention-to-scenes vs. Attention-to-objects) independent ANOVAs for the α and β parameter values. The α parameter ANOVA confirmed a significant main effect of Age, $p < .001$, indicating higher levels of attention to relevant information in adults. The other effects did not reach significance ($ps > .05$). Therefore, while we found developmental differences in focusing, these differences were not specific for stimulus type, transpiring for both objects and scenes.

We performed a similar analysis of β parameter values. The main effect of Age was significant, $p < .001$, suggesting higher levels of attention to goal-irrelevant stimuli in children compared to adults. Importantly, there was also a significant interaction between the factors, $p = .004$. Specifically, in children there were higher values of the β parameter in the attention-to-scenes condition than in the attention-to-objects condition (suggesting more attention to irrelevant objects than scenes), $t(67) = 2.29$, $p = .03$, $d = .55$, whereas no condition differences transpired in adults, $t(78) = 1.81$, $p = .08$, $d = .41$.

Table 2
ANOVA on Estimated Best-Fitting MPT Parameter Values

ANOVA	Factor	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
Age (Child vs. Adult) \times Attentional Focus	Age	(1, 145)	50.20	< .001	.26
(Attention-to-scenes vs. Attention-to-objects) \times Parameter	Attentional focus	(1, 145)	3.48	.06	.02
(α vs. β)	Parameter	(1, 145)	1,275.08	< .001	.90
	Age \times Attentional Focus	(1, 145)	0.31	.56	.00
	Age \times Parameter	(1, 145)	158.25	< .001	.52
	Attentional Focus \times Parameter	(1, 145)	0.00	.99	.00
	Age \times Attentional Focus \times Parameter	(1, 145)	9.83	.002	.06
α Parameter	Age	(1, 145)	165.51	< .001	.53
Age (Child vs. Adult) \times Attentional Focus	Attentional focus	(1, 145)	1.37	.24	.01
(Attention-to-scenes vs. Attention-to-objects)	Age \times Attentional Focus	(1, 145)	3.03	.08	.02
β Parameter	Age	(1, 145)	22.42	< .001	.13
Age (Child vs. Adult) \times Attentional Focus	Attentional focus	(1, 145)	2.03	.16	.01
(Attention-to-scenes vs. Attention-to-objects)	Age \times Attentional Focus	(1, 145)	8.70	.004	.06

Note. Statistically significant results are displayed in bold font. ANOVA = analysis of variance; MPT = Multinomial Processing Tree.

These results suggest that the pattern of observed performance (i.e., the developmental asymmetry in processing of objects and scenes) was due to improvements in the filtering of irrelevant objects. We suggest that this developmental pattern may be indicative of object-biased attention in children, but not adults. Although the behavioral results of adults in this experiment suggested some interference stemming from scenes, these effects were small and comparable to those in children. Additionally, we did not find robust evidence of differences between the scene and object conditions in MPT model parameters. We therefore conclude that adults' attention is relatively unbiased.

Overall, the reported results suggest a developmental shift from object-biased attention to more unbiased attention. The MPT model results provide a more fine-grained estimation of the attentional bias, suggesting that it stems from children's failure to filter irrelevant objects rather than from their failure to focus on relevant scenes.

General Discussion

In this work, we investigated the development of selective attention to natural scenes and objects. To do so, we used an RSVP paradigm and instructed preschool-aged children and adults to attend to either an object or a scene in target images containing both types of stimuli. Each RSVP stream contained either a congruent test image in which the object and scene both matched or mismatched the target image, or an incongruent image in which one stimulus matched and the other mismatched the target.

Behavioral results suggested that children's performance was affected more by the presence of irrelevant objects than irrelevant scenes, suggesting a possible attentional bias toward objects. In contrast, adults' performance was relatively unbiased. For a more fine-grained analysis of the results, we implemented an MPT model that estimated attention to the task-relevant feature as well as the task-irrelevant feature in each participant. This approach allowed insight into attentional focusing and filtering. Specifically, the modeling results suggested that both the ability to focus on task-relevant information and the ability to inhibit task-irrelevant information increase with development. Crucially, children struggled to filter irrelevant object information more so than irrelevant scene information, suggesting an attentional bias toward objects. At the same time, there were no differences between

focusing on relevant objects or scenes. In contrast to children, adults' attention to relevant and irrelevant stimuli was comparable across conditions.

Developmental Changes in Attention to Scenes and Objects

These results suggest intriguing developmental changes in visual cognition, including a shift from object-biased to more unbiased selective attention. Why would children have an attentional bias toward objects? One possibility is that objects are the primary unit of attention early in development, potentially beginning in infancy (Bulf & Valenza, 2013; Spelke, 1990; Werchan et al., 2019), such that visual arrays are rapidly segmented into objects that can then be selected for further processing. If children's visual attention is primarily based on object units, then objects may attract attention more so than global scene information, even if the objects are irrelevant to the task, perhaps because features within objects are attended more easily than features distributed across different objects (Duncan, 1984; Egly et al., 1994).

It is also possible that language plays a role, such that object labels are more ubiquitous in children's vocabulary than scene labels, which could boost attention to that type of stimulus (see Vales & Smith, 2015). A related possibility is that children are simply more familiar with object categories than scene categories, which could facilitate attention to objects. However, a separate baseline experiment (reported in Appendix B) found no evidence of better performance in the RSVP task when only objects were presented than when only scenes were presented, which does not support the familiarity account. Future research is needed to gain a more fine-grained understanding of mechanistic sources of an early attentional bias toward objects.

Children's attentional bias toward objects could be related to a previously hypothesized bias in children's attention toward local elements of a scene, such as an individual tree in a forest scene. Global attention to the overall scene (the forest) has been shown to develop more slowly (Dukette & Stiles, 2001; Poirel, Mellet, Houdé, & Pineau, 2008). One interesting avenue for future work would be to explore the relation between local versus global biases and object versus scene biases using an individual differences approach.

Children's attentional bias toward objects could be influenced by the developmental time-courses of different brain regions. Previous work has suggested earlier development of regions supporting

object recognition (i.e., lateral occipital complex) compared to regions supporting scene recognition (i.e., parahippocampal place area; Golarai et al., 2007). One possibility, then, is that regions supporting scene processing are relatively under-developed in children, leading to a greater reliance on object processing. Future work should aim to better understand how attentional biases relate to neural development.

Selective Attention Development

A great deal of work has suggested that attentional filtering improves across childhood (Enns & Akhtar, 1989; Plebanek & Sloutsky, 2017; Plude et al., 1994; Wendelken et al., 2011). For example, one study (Day & Stone, 1980) asked 5- and 8-year-old children and adults to identify whether a test stimulus (a line drawing of an object) matched a previously presented target stimulus. On some trials, the test stimulus was presented following an RSVP stream of distractor stimuli, whereas on other trials the test stimulus was presented alone. The ability to correctly respond to the test stimulus improved with age, especially in the RSVP condition. These results imply that the ability to filter out the irrelevant images in the RSVP stream improved with age.

The results of this study extend this work (Day & Stone, 1980) in several ways. First, every test image in our task was embedded within an RSVP stream, such that a need to filter irrelevant stimuli was always present. Our task required an additional attentional demand, which was to attend only to scenes or objects, and ignore the other type of stimulus. Because scenes and objects supported different responses on incongruent trials, selective attention to the correct stimulus was essential for high performance. In addition, our results allowed insight into potential differences between focusing and filtering of scenes and objects. Crucially, the MPT results led to the conclusion that children had more difficulty inhibiting *irrelevant* objects than scenes, whereas they had no greater difficulty focusing on *relevant* scenes compared to *relevant* objects. Therefore, the attentional bias toward objects in children may stem from immature filtering. This asymmetry between focusing and filtering of object information may be related to findings suggesting that the development of selective attention may be driven more by changes in filtering compared to focusing (Sloutsky & Plebanek, under review). At the same time, we found no differences between either focusing or filtering for objects versus scenes in adults. We therefore

conclude that in the course of the development of selective attention people acquire the ability to filter irrelevant object information.

Conclusions

This work investigated the development of selective attention to natural scenes and objects. Whereas a great deal of work has suggested that children's attention is often less selective and more diffuse, which may lead to the hypothesis that children may be biased to attend more to spatially distributed features across an entire scene, children showed an attentional bias toward objects, which supports the hypothesis of object-based attention early in development (Bulf & Valenza, 2013; Spelke, 1990; Werchan et al., 2019). In contrast, adults showed relatively unbiased attention. The MPT model results provide a more fine-grained estimation of the attentional bias in children, suggesting that it stems from children's failure to filter irrelevant objects rather than from their failure to focus on relevant scenes. Overall, these results indicate important developmental changes in selective attention and visual cognition for natural scenes and objects.

References

- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57–86. <https://doi.org/10.3758/BF03210812>
- Broadbent, D. E. (1958). *Perception and communication*. Oxford, UK: Oxford University Press.
- Bulf, H., & Valenza, E. (2013). Object-based visual attention in 8-month-old infants: Evidence from an eye-tracking study. *Developmental Psychology*, 49, 1909–1918. <https://doi.org/10.1037/a0031310>
- Chen, Z. (2012). Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, 74, 784–802. <https://doi.org/10.3758/s13414-012-0322-z>
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15, 559–564. <https://doi.org/10.1111/j.0956-7976.2004.00719.x>
- Day, M. C., & Stone, C. A. (1980). Children's use of perceptual set. *Journal of Experimental Child Psychology*, 29, 428–445. [https://doi.org/10.1016/0022-0965\(80\)90105-8](https://doi.org/10.1016/0022-0965(80)90105-8)
- Deng, W. S., & Sloutsky, V. M. (2015). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, 51, 392–405. <https://doi.org/10.1037/a0038749>
- Deng, W. S., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, 91, 24–62. <https://doi.org/10.1016/j.cogpsych.2016.09.002>

- Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 448–456. <https://doi.org/10.1037/h0090403>
- Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science*, 4, 233–251. <https://doi.org/10.1111/1467-7687.00168>
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517. <https://doi.org/10.1037/0096-3445.113.4.501>
- Egely, 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, 161–177. <https://doi.org/10.1037/0096-3445.123.2.161>
- Enns, J. T., & Akhtar, N. (1989). A developmental study of filtering in visual attention. *Child Development*, 60, 1188–1199. <https://doi.org/10.2307/1130792>
- Gerson, A. D., Parra, L. C., & Sajda, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *NeuroImage*, 28, 342–353. <https://doi.org/10.1016/j.neuroimage.2005.06.026>
- Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D. E., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, 10, 512–522. <https://doi.org/10.1038/nn1865>
- Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, 13, 622–635. <https://doi.org/10.1111/j.1467-7687.2009.00921.x>
- Horn, S. S., Bayen, U. J., & Michalkiewicz, M. (2020). The development of clustering in episodic memory: A cognitive-modeling approach. *Child Development*. <https://doi.org/10.1111/cdev.13407>
- Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, 47, 3286–3297. <https://doi.org/10.1016/j.visres.2007.09.013>
- Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, 111, 880–913. <https://doi.org/10.1037/0033-295X.111.4.880>
- Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. London, UK: Sage.
- Pastò, L., & Burack, J. A. (1997). A developmental study of visual attention: Issues of filtering efficiency and focus. *Cognitive Development*, 12, 523–535. [https://doi.org/10.1016/S0885-2014\(97\)90021-6](https://doi.org/10.1016/S0885-2014(97)90021-6)
- Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, 28, 723–732. <https://doi.org/10.1177/0956797617693005>
- Plebanek, D. J., & Sloutsky, V. M. (2019). Selective attention, filtering, and the development of working memory. *Developmental Science*, 22, e12727. <https://doi.org/10.1111/desc.12727>
- Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, 86, 227–272. [https://doi.org/10.1016/0001-6918\(94\)90004-3](https://doi.org/10.1016/0001-6918(94)90004-3)
- Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: Developmental changes during childhood in the processing of visual local-global patterns according to the meaningfulness of the stimuli. *Developmental Psychology*, 44, 245–253. <https://doi.org/10.1037/0012-1649.44.1.245>
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40, 49–70. <https://doi.org/10.1023/A:1026553619983>
- Sloutsky, V. M., & Plebanek, D. J. (under review). *Filtering of task-irrelevant information drives the development of selective attention*.
- Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14, 29–56. https://doi.org/10.1207/s15516709cog1401_3
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149. <https://doi.org/10.3758/BF03207704>
- Touryan, J., Gibson, L., Horne, J. H., & Weber, P. (2011). Real-time measurement of face recognition in rapid serial visual presentation. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00042>
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411–443. <https://doi.org/10.1080/13506280500195250>
- Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, 33, 527–532. <https://doi.org/10.3758/BF03202934>
- Turner, B. M., & Sederberg, P. B. (2012). Approximate Bayesian computation with differential evolution. *Journal of Mathematical Psychology*, 56(5), 375–385. <https://doi.org/10.1016/j.jmp.2012.06.004>
- Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science*, 18, 65–79. <https://doi.org/10.1111/desc.12179>
- Vierck, E., & Miller, J. (2006). Effects of task factors on selection by color in the rapid serial visual presentation (RSVP) task. *Perception & Psychophysics*, 68, 1324–1337. <https://doi.org/10.3758/BF03193731>
- Wendelken, C., Baym, C. L., Gazzaley, A., & Bunge, S. A. (2011). Neural indices of improved attentional modulation over middle childhood. *Developmental Cognitive Neuroscience*, 1, 175–186. <https://doi.org/10.1016/j.dcn.2010.11.001>
- Werchan, D. M., Lynn, A., Kirkham, N. Z., & Amso, D. (2019). The emergence of object-based visual attention in infancy: A role for family socioeconomic status and

competing visual features. *Infancy*, 24, 752–767. <https://doi.org/10.1111/inf.12309>

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science*, 24, 2163–2172. <https://doi.org/10.1177/0956797613487385>

Appendix A

Task Instructions

Participants were first instructed that they would use the computer mouse to make responses. For children, the left mouse button had a yellow sticker, and the right button had a pink sticker, and subsequent instructions referred to the yellow and pink buttons. For adults, the mouse did not have stickers and instructions referred to as the left and right buttons. All participants were informed that they would make responses sometimes with the yellow or left button, and sometimes with the pink or right button. Child participants were instructed to hold the mouse using both hands, with one thumb on each button, whereas adults were allowed to rest their right hand on the mouse as they typically would.

Following the mouse instructions, all participants were informed that “In this game you will see lots of pictures of objects and scenes. Objects are things like cars, trees, slides, cats, and persons. Scenes are places you can visit, like beaches, kitchens, mountains, offices, and streets.” For each stimulus category, participants were shown an example image that was not included in the main experiment.

At this time, participants were first instructed to pay attention to only objects (or scenes), and ignore the other type of stimulus, throughout the task. Following this instruction, participants were shown an example of an object (a cat) near the top of the screen, followed by an example of a scene (a kitchen), which was followed in turn by an image of the object superimposed on the scene. Participants were then shown two additional object-scene dyads, one near each bottom corner of the screen, for a total of three dyads presented simultaneously on the screen. One of these bottom images had the same object but a different scene compared to the image at the top of the screen, whereas the other image had the same scene but a different object. Participants were instructed to find the picture on the bottom of the screen with the same object (or scene, depending on condition) as the picture at the top. In order to make the correct response, participants needed to focus on the target object or scene

and ignore the nontarget stimulus. Children were asked to point to the target image, whereas adults were asked to press the left or right mouse button to make a response. If children were confused or made the wrong response, the experimenter provided additional explanation.

Training Phase

Following this, participants completed a short training phase for the rapid serial visual presentation (RSVP) task. On each training trial, just as in the main experiment, participants were shown a target object and scene separately (in random order), followed by a combined image of the object superimposed on the scene. Participants were then presented with an RSVP stream as in the main task.

The training phase of the experiment included six trials. The instructions accompanying the trials increased in complexity. On the first two trials, following the presentation of the target stimuli, participants were told they would see other objects and scenes that would “go by fast,” and that their job was to watch for exactly the same object (or scene) as they saw on the combined target stimulus, and ignore the other type of stimulus. After the completion of the stream, they were asked whether they saw the target object (or scene). If they had, they were asked to press the appropriate button on the mouse to signify that the stimulus was repeated, or if they had not, they were asked to press the other button. On the third and fourth trials, participants were asked to watch for the same object (or scene), and to press the “old” response button on the mouse immediately if they saw it. If participants had not responded by the end of the stream, participants saw a question mark on the screen and were asked whether they had seen it, and to press either the “old” or “new” response button as appropriate. On the final two training trials, participants were instructed prior to the start of the RSVP stream that they should press the “old” response button immediately if they saw a repetition of the relevant target stimulus, or press the “new” response button at the end of the trial if they did not see the same stimulus.

Following the response made on every training trial, participants were provided with feedback, such as “Great job! The same object [or scene] WAS there that time!” or “Uh oh! The same object [or scene] was NOT there that time!” Correct and incorrect responses were accompanied by a smiling or frowning face, respectively. If a response was

made before the target stimulus was presented, participants were told “Uh oh! You pressed a button too soon!” and saw a frowning face.

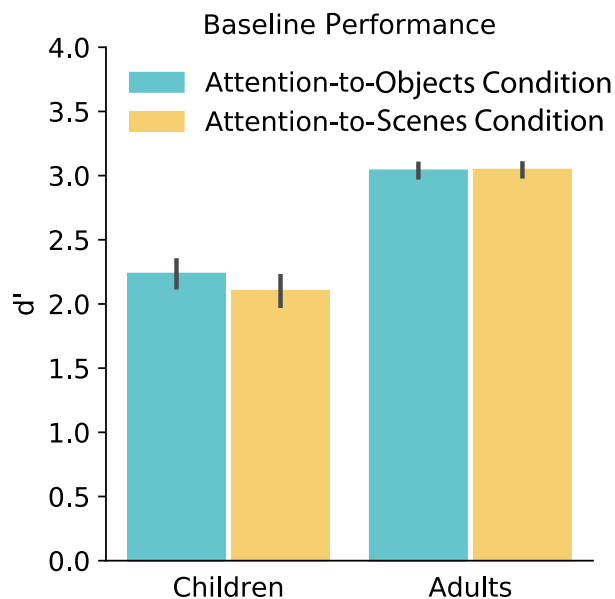
If children asked questions or were confused at any point in the training phase, the experimenter elaborated on the instructions as needed. The order of correct responses across these six trials for each participant was “old,” “old,” “old,” “new,” “old,” “new.” In order to encourage participants to focus on the task-relevant and ignore the irrelevant stimuli, these two stimuli always supported incongruent responses (i.e., when the relevant feature was old, the irrelevant feature was new, and vice versa).

Upon completion of the training phase, participants were told that they would continue playing the same game, but now they would not receive feedback on their responses. They were also briefly reminded of which button to press immediately when the target stimulus was repeated during the RSVP stream, and which button to press at the end when it was not repeated. Participants were again briefly reminded of these instructions at the start of each block of trials.

Appendix B

Baseline Performance

In a separate experiment conducted between February 2017 and October 2017, we measured children’s ($N = 60$; $M_{\text{age}} = 5.7$ years, $SD = .5$, range = 5.0–6.9; 28 females, 32 males) and adults’ ($N = 75$; 53 females, 22 males) ability to perform the RSVP task

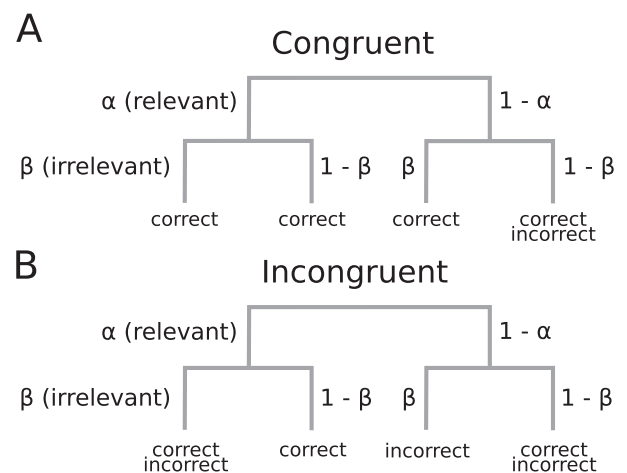


without the distracting presence of goal-irrelevant stimuli. In this baseline condition, participants were asked to attend to objects (or scenes) in the absence of the other stream of information for one block of 20 trials. Participants’ performance on this baseline task is presented in Figure B1. To assess potential effects of age and stimulus condition, we conducted a 2 (Age: Child vs. Adult) \times 2 (Stimulus: Objects vs. Scenes) between-subjects analysis of variance. There was a significant main effect of Age, $F(1, 131) = 118.39$, $p < .001$, $\eta_p^2 = .47$, indicating higher performance in adults. There was no effect of the Stimulus condition, $p = .53$, and no interaction, $p = .40$. Because performance on this baseline task did not differ between objects and scenes, we suggest that the children’s asymmetric differences between congruent and incongruent trials in the main experiment were unlikely to be due to differences in familiarity with these stimulus types.

Appendix C

MPT Model

Multinomial processing tree (MPT) models are based on tree-like structures that define sequences of latent events. The MPT structures for the congruent and incongruent conditions of the current experiment are presented in Figure C1. Attention to the relevant and irrelevant stimuli was estimated by parameters α and β , respectively. The “branches” of the tree structure define how different sequences of processes are assumed to result in different outcomes; in some cases, multiple branches result in the same outcome. For example, on congruent trials, the correct response can be the result of any of the four branches, whereas an incorrect response can only be



the result of a lack of attention to both the congruent and incongruent stimulus. Mathematically, the probability of each branch is defined as the product of the parameter values making up that branch, and the probability of each type of response is the sum of the probabilities defined by all branches that could result in that response. For example, for congruent trials, the probability of a correct response is $(\alpha \times \beta) + (\alpha \times (1 - \beta)) + ((1 - \alpha) \times \beta) + ((1 - \alpha) \times (1 - \beta) \times 0.5)$. Note that the last term is multiplied by 0.5 because without attention to the relevant or irrelevant stimulus $((1 - \alpha) \times (1 - \beta))$, we assume responses will be the result of guessing equally between the correct and incorrect responses.

The model equations used to estimate responses in each condition of the experiment are presented below in Python code.

```
# guess 'new' or 'old' with equiprobability
guess = 0.5
# rel = relevant stimulus; irrel = irrelevant
stimulus
# alpha and beta = parameter values between 0 and 1
# theta = list of response probabilities for every
trial type
# rel new, irrel new, respond new
theta[0] = alphabeta + alpha(1-beta) + (1-alpha)
beta + (1-alpha)(1-beta)guess
# rel new, irrel new, respond old
theta[1] = (1-alpha)(1-beta)guess
# rel new, irrel old, respond new
theta[2] = alphabeta + alpha(1-beta) + (1-
alpha)(1-beta)guess
# rel new, irrel old, respond old
theta[3] = alphabeta + (1-alpha)beta + (1-
alpha)(1-beta)guess
# rel old, irrel new, respond new
theta[4] = alphabeta + (1-alpha)beta + (1-
alpha)(1-beta)guess
# rel old, irrel new, respond old
theta[5] = alphabeta + alpha(1-beta) + (1-
alpha)(1-beta)guess
```

```
# rel old, irrel old, respond new
theta[6] = (1-alpha)(1-beta)guess
# rel old, irrel old, respond old
theta[7] = alphabeta + alpha(1-beta) + (1-alpha)
beta + (1-alpha)(1-beta)guess
```

Bayesian Model fitting

We fitted our MPT model to the data of each participant independently using a Bayesian approach. To implement our analyses, we applied custom programs with RunDEMC (<https://github.com/compmem/RunDEMC>), which uses a differential evolution algorithm to produce Markov Chain Monte Carlo (MCMC) simulations (Turner & Sederberg, 2012).

The model had two free parameters that estimated attention to task-relevant and task-irrelevant stimuli, as described in the main text. To estimate the model parameters in a Bayesian framework, we specified the following uniform prior distributions:

$$\alpha \sim u(0,1),$$

$$\beta \sim u(0,1).$$

We fitted the model using 1,000 iterations of the MCMC algorithm with 20 chains. The first 200 iterations of each chain were used as a burn-in period and discarded from the analysis. Best-fitting parameter values for each participant were calculated using maximum a posteriori estimates. These best-fitting parameter values for each participant were the basis of the statistical analyses presented in the main text. We also used these parameter values to generate the model-predicted proportion of each response in each condition according to the equations summarized above. These model-generated data are presented in Figure 2A of the main text.