

# Children’s use of local and global visual features for material perception

**Benjamin Balas**

Psychology Department, North Dakota State University,  
Fargo, ND, USA

**Amanda Auen**

Psychology Department, North Dakota State University,  
Fargo, ND, USA

**Josselyn Thrash**

Psychology Department, North Dakota State University,  
Fargo, ND, USA

**Shea Lammers**

Department of Human Development and Family Science,  
North Dakota State University, Fargo, ND, USA

Adults can rapidly recognize material properties in natural images, and children’s performance in material categorization tasks suggests that this ability develops slowly during childhood. In the current study, we further examined the information children use to recognize materials during development by asking how the use of local versus global visual features for material perception changes in middle childhood. We recruited adults and 5- to 10-year-old children for three experiments that required participants to distinguish between shape-matched images of real and artificial food. Accurate performance in this task requires participants to distinguish between a wide range of material properties characteristic of each category, thus testing material perception abilities broadly. In two tasks, we applied distinct methods of image scrambling (block scrambling and diffeomorphic scrambling) to parametrically disrupt global appearance while preserving features in small spatial neighborhoods. In the third task, we used image blurring to parametrically disrupt local feature visibility. Our key question was whether or not participant age affected performance differently when local versus global appearance was disrupted. We found that although image blur led to disproportionately poorer performance in young children, this effect was reduced or absent when diffeomorphic scrambling was used. We interpret this outcome as evidence that the ability to recruit large-scale visual features for material perception may develop slowly during middle childhood.

## Introduction

Material perception, the ability to categorize objects and surfaces based on what they are made of, is a critically important aspect of high-level vision (Fleming, 2013). Material judgments support inferences regarding the expected tactile properties of objects (Baumgartner, Wiebel, & Gegenfurtner, 2013), the effects of various transformations of shape (e.g., ripping, folding; Schmidt & Fleming, 2018), the freshness or ripeness of food (Arce-Lopera et al., 2012), and many other variables. Adults are capable of estimating material properties given brief exposure to complex images (Sharan, Rosenholtz, & Adelson, 2009, 2014; Wiebel, Valsecchi, & Gegenfurtner, 2014), and sensitivity to material properties emerges early in neural responses (Baumgartner & Gegenfurtner, 2016; Jacobs, Baumgartner, & Gegenfurtner, 2014), suggesting that such judgments are carried out rapidly and effectively even when visual information is impoverished (Balas, Conlin, & Shipman, 2017). Although adults’ abilities to categorize and estimate material properties from natural images have been examined in many studies, there are still little data describing the developmental trajectory of material perception. In the current study, we examined how children and adults use visual information across different spatial scales to estimate material properties. Specifically, we were interested in determining the extent to which children between 5 and 10 years of age and adults tend to rely on small-scale, local visual features versus larger-scale, global visual features to estimate material properties.

Citation: Balas, B., Auen, A., Thrash, J., & Lammers, S. (2020). Children’s use of local and global visual features for material perception. *Journal of Vision*, 20(2):10, 1–13, <https://doi.org/10.1167/jov.20.2.10>.



Middle childhood, in particular the target age range described above, is a particularly interesting developmental stage to examine in the context of material perception for two main reasons. First, texture perception (which we assume makes a substantial contribution to material perception) appears to develop slowly during these years. Children do not exhibit adult-like texture discrimination abilities until at least 10 years of age (Elleberg, Hansen, & Johnson, 2012) and also do not appear to segment textures accurately until late in middle childhood (Sirteneau & Rieth, 1992). In previous work, we have also shown that regarding material perception specifically, some aspects of material categorization are also not adult-like during middle childhood, although this varies by material property (Balas, 2017). A key question regarding this slow development of adult-like abilities is whether or not ongoing improvements in material perception reflect changes in how visual information is used to inform material judgments or more general improvement in the efficiency of processes that support material perception. In our previous work, we found that forcing children and adults to use visual summary statistics (Balas, 2006; Balas, Nakano, & Rosenholtz, 2009) to assess material properties affected categorization, but not matching, performance differently as a function of age (Balas, 2017). We interpreted these results as evidence that children had more or less adult-like abilities to compare and match materials using the information available in “mongrels” made from natural material images but had not yet established robust mappings between those features and category labels. More broadly, we think this is consistent with an account of the development of material perception in which children do not rely on different features for material judgments but may use the same features as adults less effectively. An important limitation of this study, however, is the fact that we only examined children’s and adults’ performance subject to one manipulation of appearance: the application of a texture synthesis model. The issue here is that using any one model to manipulate the visual information available for material perception is that the model constrains what features can and cannot be disrupted. In the case of the Portilla-Simoncelli model (Portilla & Simoncelli, 2000) applied in our previous report, a number of joint wavelet statistics within local image neighborhoods tend to be preserved, whereas higher-order statistics within those neighborhoods and relationships at scales larger than those neighborhoods are generally not preserved. As a result, using this model (or similar pyramid-based models; Briand et al., 2014; Heeger & Bergen, 1995) to study visual development represents a commitment to examining the impact of disrupting features at large spatial scales more than disrupting features at smaller spatial scales.

Considering the visual information available at different spatial scales may be critically important to understanding how material perception develops. To the extent that texture features are important for estimating material properties, a range of cues are available across relatively small and relatively large spatial scales—the visual perception of roughness, for example, depends on high spatial frequency information in some observers and low spatial frequency in others (Bergmann & Kappers, 2007). A number of other material properties, such as gloss, are visually judged based on local features but also on larger-scale cues to surface shape (Anderson & Kim, 2009). Children and adults thus have the opportunity (and in some cases the need) to recruit visual features across multiple scales to estimate material properties accurately. However, in other problem domains, children appear to rely more heavily on local features early in childhood, eventually relying on global information (or at information at larger spatial scales) relatively late in childhood. This developmental trajectory has frequently been established via the use of Navon-like hierarchical figures in which small shapes (e.g., letters) are arranged to make a large shape that may or may not differ from that of the constituent elements (a large T made up of small xs). Relative to adults, young children exhibit superior abilities to report the shape of local elements in such figures as measured via drawing tasks (Dukette & Stiles, 2001), discrimination and search tasks (Kimchi et al., 2005), and eye movements (Vurpillot, 1968). The bias for local features continues into adolescence (Scherf et al., 2009) and may depend on structural changes in the developing brain (Poirel et al., 2011). These results suggest that although there may be useful information available for material perception at large spatial scales, young children may not use that information as readily as adults, leading to different patterns of performance as visual features at local and global spatial scales are selectively disrupted. To our knowledge, however, there have been few studies in which the development of local and global visual processing has been examined using complex natural images rather than structured, schematic stimuli.

Our goal was to test the hypothesis that children’s material perception abilities may depend disproportionately on local (small-scale) visual features relative to adults. To examine this question, we asked children and adults to distinguish between images of real food and images of fake food (plastic, fabric, or wooden versions of real foods). We selected this task for several reasons. In terms of material perception, distinguishing real foods from artificial foods requires that observers be capable of estimating many different material properties, including glossiness, hardness, and roughness. Using a real/artificial food judgment thus provides a means of studying material perception broadly as opposed to focusing on a specific material

property. We also chose this task because we wished to use a judgment that both young and older children would find intuitive. Given that children of all ages have extensive experience with real food and are likely to have encountered artificial food in play settings, this task was an attractive option. We are not arguing that real/artificial material judgments of this kind are particularly useful for examining local and global processing but rather that this task was a useful vehicle for examining these questions for the reasons we described above. In separate experiments, we asked participants to perform this task subject to image manipulations that either disrupted local visual features more than global (or large-scale) visual features via image blur ([Experiment 1](#)) or disrupted global visual features more than local ones via image scrambling ([Experiments 2 and 3](#)). We reasoned that if children indeed showed a greater bias toward local visual features than adults, making those features less available should impair performance disproportionately for our youngest participants. Critically, manipulations that tended to preserve local visual features should not disproportionately affect child participants, as the information they rely on most heavily should still be available. Briefly, we found that limiting the availability of local image features did impair young children's performance more than adults but that weaker trends in this direction were also evident when image scrambling was imposed. We discuss these results in the context of other results examining how children recruit information across spatial scales for other recognition tasks and more broadly in terms of how visual integration abilities may develop generally during childhood.

## Experiment 1

In our first experiment, we examined how removing local visual cues to material properties affected material perception in children and adults. Specifically, we used varying amounts of image blur to limit the availability of fine details that children and adults could use to classify material properties.

## Methods

### Participants

Our final sample comprised a total of 62 participants, including children 5 to 7 years old ( $n = 21$ , 13 female), children 8 to 10 years old ( $n = 22$ , 13 female), and adult participants between the ages of 18 and 25 years ( $n = 19$ , 11 female). Child participants were recruited from the Fargo-Moorhead community, and adult participants were recruited using the North Dakota State University (NDSU) Undergraduate Psychology

Study Pool. All participants self-reported normal or corrected-to-normal acuity and normal color vision, and they were naive to the purpose of the study and the main hypotheses. Adult participants received course credit for their participation, whereas child participants received monetary compensation and their choice of a book to take home. We obtained written informed consent from all adult participants and the parent/guardian accompanying each child participant. Children also provided written assent to participate prior to the beginning of the experiment session.

### Stimuli

Our stimuli were created using 96 full-color images of real and artificial food obtained via Google Images. We selected 48 image pairs such that each image of a real foodstuff was paired with a matching image of an artificial item. To the extent possible, these pairs were selected so that the shape, orientation, color, and other appearance variables were approximately matched within each pair ([Figure 1](#)). The original images varied in size, so we cropped and resized all images to  $512 \times 512$  pixels and imposed a uniform white background. These original images served as the basis for all three of the experiments described here. Across images of real and artificial food, material properties varied substantially. Artificial foods included objects made of plastic, wood, and fabric, for example, whereas real foods varied in glossiness, hardness, roughness, and other material properties. We made no attempt to balance these material categories and properties either within or between real and artificial food categories, and so our stimulus set does not support a detailed examination of how specific material properties influenced performance. Instead, these stimuli were intended to discourage the use of a particular visual cue to support real/artificial discrimination while also requiring participants to consider a broad set of material properties to achieve high levels of accuracy. Finally, we note that in general, the foods we used in these tasks were items that we expected children in our target age range to be able to recognize. Using the [Kuperman et al. \(2012\)](#) Age-of-Acquisition norms, we found that the average age at which children knew the words for our food items was approximately 4.9 years ( $SD = 1.47$ ). As such, we are confident that children in all three studies were likely to be sufficiently familiar with these foods that they could name them and therefore not struggle with the chance due to a lack of prior experience with the food items depicted here.

For this experiment, we created additional stimuli by using image blur to parametrically vary the amount of fine spatial detail available to our observers for material perception. We used the `blur.m` function implemented in the Steerable Pyramid Toolbox for MATLAB to create new images with low levels of blur (two levels of



Figure 1. Example stimuli created by blurring. This artificial avocado would be paired with its counterpart real avocado during the task.

filtering and downsampling, followed by upsampling to the original image size) and high levels of blur (four levels rather than two). In both cases, we used a binomial filter kernel (the default “binom5” kernel in `blur.m`). The resulting images remained  $512 \times 512$  pixels in size and were full-color (Figure 1). The full set of stimuli used in all of our experiments is available via the Open Science Framework (OSF) at the following link: <https://osf.io/hvmpn/>.

## Procedure

### Real/artificial food task

We asked our participants to complete a 2-alternative forced choice (2AFC) food/nonfood task using the stimuli described above. On each trial of this task, participants were presented with a real/artificial image pair (e.g., a real avocado and its matching plush counterpart) displayed to the left and right of the screen. Participants were asked to indicate which of these two images depicted a real food using two large response buttons placed near their left and right hands. The images remained onscreen until participants made a response, but participants were asked to respond as quickly as possible while being accurate. Specifically, we instructed all participants as follows: “Please tell us which of the two pictures is a real food that someone could eat. To choose the left picture, press the left button, and to choose the right picture, press the right button. Please try to go as quickly as you can while being careful to pick the right picture!” The left/right position of real food images varied pseudorandomly across trials. Image blur also varied pseudorandomly across trials, and every image pair appeared once at each level of blur (no blur, low blur, and high blur) for a grand total of 144 trials in the entire task. Participants did not receive feedback regarding accuracy during the task.

Participants completed the task seated approximately 40 cm away from a MacBook Pro laptop with a 1,200-pixel  $\times$  800-pixel display. Food images were

presented offset to the left and right of center by 300 pixels and subtended approximately 4 to 5 degrees of visual angle. All stimulus display and response collection routines were executed using custom routines written using the Psychtoolbox v3.0 extensions for MATLAB (Kleiner et al. 2007).

### Baseline response latency task

Because we were interested in comparing response latencies across child and adult participants, we included an additional task designed to provide a baseline measure of response latency differences across these participant groups. Before completing the real/artificial food task described above, we asked each participant to complete a short 2AFC color categorization task using the same experimental setup. Specifically, on each trial of this task, we presented participants with a red circle and a green circle, offset to the left and right of center, with the left/right position of the two circles pseudorandomly varied across trials. Participants were asked to use the two response buttons to indicate where the red circle was on each of 32 trials. As in the task described above, we asked participants to respond as quickly as possible while remaining accurate.

## Results

We analyzed participants’ accuracy and response latency across blur levels as a function of age group using a  $3 \times 3$  mixed analysis of variance (ANOVA) implemented in JASP (JASP Team, 2018). All aggregate data files and raw MATLAB files from each participant are available via OSF at the following link: <https://osf.io/hvmpn/>.

### Accuracy

For each participant, we calculated the proportion of correct trials per blur condition. The results of an ANOVA run using these values yielded main effects of blur level [ $F(2, 118) = 101.7, p < 0.001$ , partial

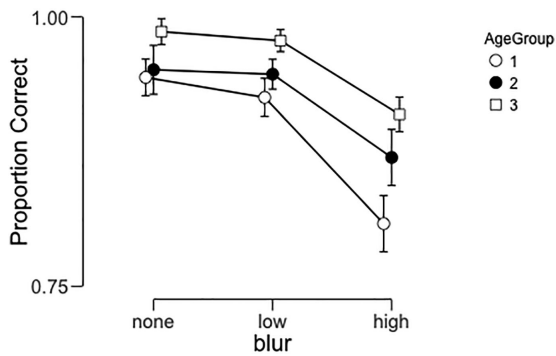


Figure 2. Average accuracy as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and blur level (none, low, and high). Error bars indicate 95% confidence intervals.

$\eta^2 = 0.63$ ] and age group [ $F(2, 59) = 7.5, p = 0.001$ , partial  $\eta^2 = 0.20$ ]. We examined both of these main effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. The main effect of blur level was driven by significantly lower accuracy for high blur images relative to low blur images ( $t = 12.98$ , Cohen's  $d = 1.65, p < 0.001$ ) and unblurred images ( $t = 11.61$ , Cohen's  $d = 1.47, p < 0.001$ ). The main effect of age was driven by significant differences between 5- to 7-year-olds and adults ( $t = 3.88$ , Cohen's  $d = 0.49, p < 0.001$ ). Besides these main effects, we also observed a significant interaction between blur level and age group,  $F(4, 118) = 3.53, p = 0.009$ , partial  $\eta^2 = 0.11$ . Upon inspection, this result appeared to be driven by the disproportionately poor performance of 5- to 7-year-old children in response to high-blur images. To investigate this further, we carried out a post hoc analysis of the size of the blur effect across age groups: We calculated a difference score by subtracting performance in the “high-blur” condition from performance in the “low-blur” condition for each participant and then analyzed those difference scores using a one-way ANOVA with age group as a between-subjects factor. This yielded a significant effect of age group,  $F(2, 59) = 3.59, p = 0.034$ , which further post hoc testing revealed was the result of significant differences in the size of this difference score between adults and young children (post hoc Tukey's test,  $t = 2.48, p = 0.042$ ). This suggests that increased image blur hurts performance in general but appears to hurt 5- to 7-year-olds more than adults. In Figure 2, we include a plot of average accuracy as a function of age and blur level.

### Response latency

First, we analyzed the response time to our baseline motor task across all age groups. We carried out a

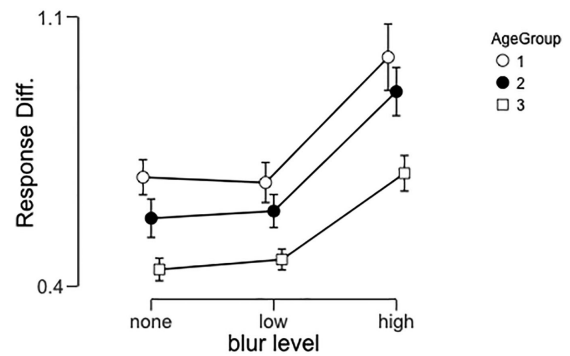


Figure 3. Average median response latency as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and blur level (none, low, and high). Error bars indicate 95% confidence intervals.

one-way ANOVA with age group as a between-subjects factor. This analysis revealed a significant effect of age group,  $F(2, 59) = 45.2, p < 0.001$ , such that all age groups differed from one another in their response latency to correct trials (adults,  $M = 0.34, SEM = 0.034$ ; 5- to 7-year-olds,  $M = 0.78, SEM = 0.032$ ; 8- to 10-year-olds,  $M = 0.50, SEM = 0.031$ ).

Second, to calculate material perception response latency for each participant, we calculated the median response time to correct trials within each experimental condition of the real/artificial food task. Next, to account for age-related differences in overall response latency, we calculated the difference between this value and the median response latency to correct trials in the baseline response time (RT) task. The resulting difference score thus reflects a participant-specific latency to estimate material properties from the experimental stimuli, correcting for individual differences in motor responses (see Meissner et al., 2018, for a similar approach).

Our analysis of these values also yielded main effects of image blur [ $F(2, 118) = 132.0, p < 0.001$ , partial  $\eta^2 = 0.69$ ] and age group [ $F(2, 59) = 46.77, p < 0.001$ , partial  $\eta^2 = 0.61$ ]. We examined both of these main effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. The effect of blur level was the result of significantly slower latencies in response to high-blur images relative to low-blur images ( $t = 12.0$ , Cohen's  $d = 1.53, p < 0.001$ ) and unblurred images ( $t = 12.46$ , Cohen's  $d = 1.58, p < 0.001$ ). The main effect of age group was the result of significant differences between all three age groups ( $t > 4.5$  for each test, Cohen's  $d > 0.5$  in each case). Finally, the interaction between blur level and age group did not reach significance,  $F(4, 118) = 1.29, p = 0.28$ , partial  $\eta^2 = 0.042$ . In Figure 3, we include a plot average median response latency as a function of blur level and age group.

## Discussion

Our results demonstrate that limiting the amount of local information available in complex images has larger effects on younger children than on older children and adults. Accuracy appears to be disproportionately affected by high levels of image blur in 5- to 7-year-old participants, and task-specific response latency decreases gradually during middle childhood. We suggest that both of these results are consistent with the hypothesis that young children largely rely on local information for recognition tasks and develop adult-like abilities to recruit global features gradually during middle childhood. By themselves, however, these data are not sufficient to rule out a number of important alternative accounts. For example, it could be the case that young children are generally affected more by most forms of image degradation, regardless of the class of visual features that are disrupted. Were this the case, image manipulations that disrupt global information more than local features should have essentially the same impact on performance. We continue, therefore, by considering how such manipulations affect performance in our real/artificial food task.

## Experiment 2

In our second experiment, we carried out the same real/artificial food task described in [Experiment 1](#) subject to block scrambling of our stimuli. By parametrically varying the size of the blocks used to scramble our base images, we can preserve information within local neighborhoods while dramatically disrupting visual information spanning image regions larger than the block size. If young children are indeed more reliant on local image features to carry out material judgments in this task, we would expect that this manipulation should not lead to disproportionately poorer performance in child participants. However, if the key issue is simply that children struggle disproportionately with impoverished images, we would expect to observe the same type of interaction that we observed in [Experiment 1](#).

## Methods

### Subjects

We recruited a total of 65 participants for this experiment, including children 5 to 7 years old ( $n = 23$ , 15 female), children 8 to 10 years old ( $n = 22$ , 14 female), and adult participants between the ages of 18 and 25 years ( $n = 20$ , 13 female). All recruitment procedures (compensation, etc.) were the same as those described in [Experiment 1](#).

### Stimuli

The same stimuli described above in [Experiment 1](#) were used as the basis for the images employed here. In this experiment, however, we implemented a simple form of block-based scrambling to preserve information in local image neighborhoods while disrupting visual information spanning regions larger than that spatial scale. Specifically, each original image was segmented into nonoverlapping tiles of varying size ( $128 \times 128$  pixels for low scrambling,  $64 \times 64$  pixels for high scrambling), and these tiles were then randomly rearranged within a square grid. The resulting images were still  $512 \times 512$  pixels in size and in full color ([Figure 4](#)).

### Procedure

The testing procedure for this task was essentially the same as that described for [Experiment 1](#), with varying levels of scrambling rather than varying levels of blur. We also included the baseline response latency task described in [Experiment 1](#), following the same procedures outlined previously.

## Results

As described in [Experiment 1](#), we calculated participant accuracy and response latency to correct trials. We analyzed these values using separate  $3 \times 3$  mixed-design ANOVAs with scrambling level (none, low, high) as a within-subjects factor and age group (5–7 years, 8–10 years, adults) as a between-subjects factor.

### Accuracy

Our analysis of the accuracy data from subjects in this task revealed significant main effects of scrambling level [ $F(2, 124) = 11.53, p < 0.001$ , partial  $\eta^2 = 0.16$ ] and age group [ $F(2, 62) = 6.99, p = 0.002$ , partial  $\eta^2 = 0.18$ ]. We examined both of these main effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. These post hoc tests revealed that the former effect was due to significantly lower accuracy in the high scrambling condition relative to the unscrambled ( $t = 3.82$ , Cohen's  $d = 0.47, p < 0.001$ ) and low scrambling conditions ( $t = 3.57$ , Cohen's  $d = 0.44, p = 0.002$ ). The main effect of age group was due to significantly lower accuracy between 5- to 7-year-olds and adults ( $t = 3.66$ , Cohen's  $d = 0.45, p = 0.002$ ), with other comparisons not reaching significance. In contrast to the results observed in [Experiment 1](#), we did not observe a significant interaction between scrambling level and age group,  $F(4, 124) = 1.86, p = 0.12$ , partial  $\eta^2 = 0.057$ . In [Figure 5](#), we display the average proportion correct as a function of age group and scrambling level.



Figure 4. Example stimuli created via block scrambling. Visual features within tiles are preserved, but larger-scale features are disrupted.

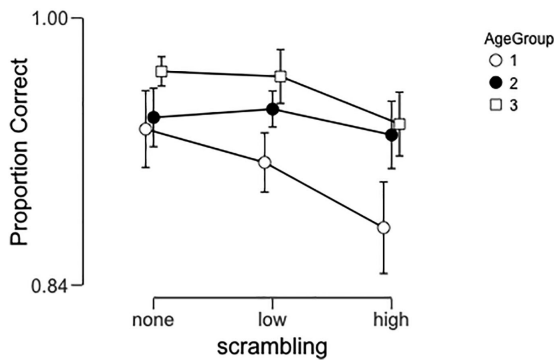


Figure 5. Average accuracy as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and scrambling level (none, low, and high). Error bars indicate 95% confidence intervals.

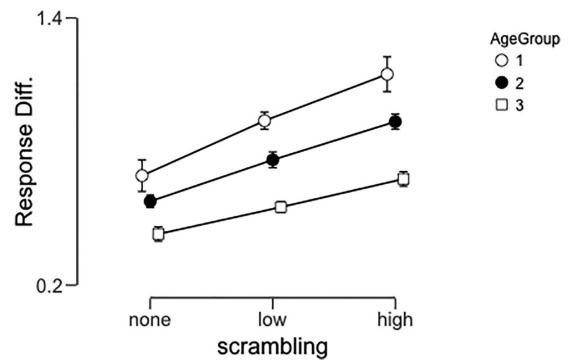


Figure 6. Average response latency as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and scrambling level (none, low, and high). Error bars indicate 95% confidence intervals.

### Response latency

First, we analyzed the response time to our baseline motor task across all age groups. We carried out a one-way ANOVA with age group as a between-subjects factor. This analysis revealed a significant effect of age group,  $F(2, 62) = 74.3$ ,  $p < 0.001$ , such that all age groups differed from one another in their response latency to correct trials, according to post hoc two-tailed  $t$  tests implemented using JASP. (In all comparisons, Bonferroni-corrected  $p$  values were less than 0.001; mean values for each group were as follows: adults,  $M = 0.32$ ,  $SEM = 0.037$ ; 5- to 7-year-olds,  $M = 0.77$ ,  $SEM = 0.036$ ; 8- to 10-year-olds,  $M = 0.44$ ,  $SEM = 0.037$ ).

Our analysis of the difference in response latencies between real/artificial categorization and our baseline measure of response latency revealed main effects of scrambling level [ $F(2, 124) = 182.7$ ,  $p < 0.001$ , partial  $\eta^2 = 0.75$ ] and age group [ $F(2, 62) = 17.9$ ,  $p < 0.001$ , partial  $\eta^2 = 0.37$ ]. We examined both of these main

effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. In both cases, these effects were the result of significant differences between all stimulus levels (Bonferroni-corrected  $p$  values were all less than 0.01). These main effects were qualified by a significant interaction between scrambling level and age group,  $F(4, 124) = 5.3$ ,  $p < 0.001$ , partial  $\eta^2 = 0.15$ . This interaction appeared to be the result of an increasing effect of age group as scrambling level increased (Figure 6), suggesting that children tended to take increasingly longer to generate correct responses as the scrambling level increased. As in Experiment 1, we investigated this interaction further via a post hoc analysis based on difference scores calculated across scrambling levels for each age group. In this case, we subtracted the scores from the “high scrambling” condition from the scores of the “no scrambling” condition to estimate the slope of the RT curves across scrambling levels. Next, we analyzed these values using a one-way ANOVA with age group as a between-subjects factor. This yielded

a significant effect of age group,  $F(2, 62) = 5.93$ ,  $p = 0.004$ , which we examined further using post hoc Tukey's tests. This revealed that the size of the difference score differed significantly between young children and adults ( $t = 3.44$ ,  $p = 0.003$ ), suggesting that young children indeed experienced a larger cost for increased scrambling relative to adults.

## Discussion

The results of [Experiment 2](#) are similar in some regards to those obtained in [Experiment 1](#) but differ in other ways. Although we observed a similar interaction between scrambling level and age group in this task, this critical interaction was evident in our response latency data rather than our accuracy data. This could mean that disrupting global information rather than local features impacts the efficiency of visual processing in this task rather than its outcome, which in itself is a potentially interesting conclusion. Alternatively, we could also interpret this result in terms of the hypothesis we advanced following [Experiment 1](#): Young children may simply find any form of image manipulation disproportionately difficult to deal with. In turn, this could mean that young children rely on a broad set of features for material perception and struggle more than adults when any particular class of features is unavailable.

One concern, however, regarding this method of disrupting large-scale visual features while preserving local information is that block scrambling as we have implemented it here introduces artifactual contours into the image. Specifically, the gridlines corresponding to the boundaries of the shuffled image blocks are visible after scrambling, which is a source of noise spanning both low and high spatial frequencies. To help address this concern, we continue in [Experiment 3](#) by examining children's and adults' performance in this task subject to a different type of image scrambling that does not introduce comparable image artifacts.

## Experiment 3

In our third and final experiment, we continued to examine how material perception in our real/artificial food task was affected by an image scrambling manipulation that disrupts large-scale visual features more than small-scale, local visual features. Specifically, we implemented a form of image scrambling called *diffeomorphic scrambling* ([Stojanowski & Cusack, 2014](#)) that has the benefit of producing images that are free from spatial frequency artifacts and preserve a larger class of image properties than simple block scrambling. This includes response properties in several layers of the HMAX model (Riesenhuber & Poggio,

2002), which is an established model of hierarchical visual processing in the human brain. Diffeomorphic scrambling thus represents a means of imposing a scrambling manipulation on complex images without introducing systematic image artifacts.

## Methods

### Subjects

We recruited a total of 61 participants for this experiment, including children 5 to 7 years old ( $n = 21$ , 14 female), children 8 to 10 years old ( $n = 19$ , 11 female), and adult participants between the ages of 18 and 25 years ( $n = 21$ , 13 female). All recruitment procedures (compensation, etc.) were the same as those described in [Experiment 1](#).

### Stimuli

The same stimuli described above in [Experiment 1](#) were once again used as the basis for the images employed here. In this experiment, however, we used diffeomorphic scrambling to produce scrambled versions of the original stimuli. This technique involves a local warping operation that is applied to the images, preserving a number of low- and mid-level features of the original image. We created our images by imposing a maximum distortion value of 80 and selected three levels of scrambling from the resulting continuum of images spanning unscrambled to maximally scrambled appearance. We selected these levels of scrambling based on the data reported in [Stojanowski and Cusack \(2014\)](#), with the goal of selecting high levels of scrambling that did not completely compromise the ability to recover meaning from the images. The resulting images were still  $512 \times 512$  pixels in size and in full color ([Figure 7](#)).

### Procedure

The testing procedure for this task was again the same as that described for [Experiment 1](#), with varying levels of scrambling rather than varying levels of blur. We also included the baseline response latency task described in [Experiment 1](#), following the same procedures outlined previously.

## Results

As in both [Experiments 1](#) and [2](#), we analyzed both accuracy and response latency data using a  $3 \times 3$  mixed-design ANOVA with scrambling level (none, low, and high) as a within-subjects factor and age group (5–7, 8–10, and adults) as a between-subjects factor.





Figure 7. Example stimuli created via diffeomorphic scrambling. Visual features within tiles are preserved, but larger-scale features are disrupted.

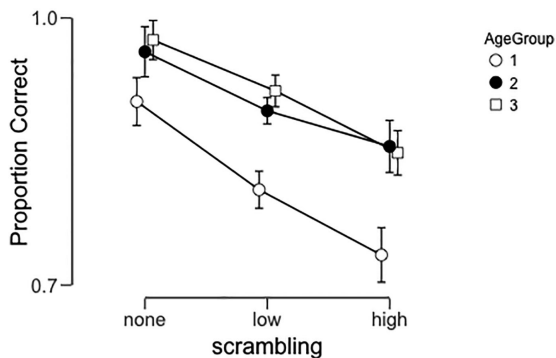


Figure 8. Average accuracy as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and scrambling level (none, low, and high). Error bars indicate 95% confidence intervals.

### Accuracy

Our analysis of participants' proportion correct responses across conditions revealed main effects of scrambling level [ $F(2, 116) = 100.1, p < 0.001$ , partial  $\eta^2 = 0.63$ ] and age group [ $F(4, 116) = 12.1, p < 0.001$ , partial  $\eta^2 = 0.30$ ]. We examined both of these main effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. The main effect of scrambling level was the result of significant differences between all three levels of scrambling, whereas the main effect of age group was the result of significantly lower accuracy in 5- to 7-year-olds relative to older children ( $t = 3.96$ , Cohen's  $d = 0.51, p < 0.001$ ) and to adults ( $t = 4.49$ , Cohen's  $d = 0.58, p < 0.001$ ). The interaction between scrambling level and age group was marginal,  $F(4, 116) = 2.38, p = 0.055$ , partial  $\eta^2 = 0.076$ , which appears to reflect a trend for disproportionately lower accuracy values in younger participants as scrambling level increases (Figure 8).

### Response latency

First, we analyzed the response time to our baseline motor task across all age groups. We carried out a one-way ANOVA with age group as a between-subjects factor. This analysis revealed a significant effect of age group,  $F(2, 58) = 34.4, p < 0.001$ , such that all age groups differed from one another in their response latency to correct trials. (Bonferroni-corrected post hoc  $t$  tests implemented in JASP yielded  $p$  values less than 0.01 for all pairwise comparisons. Mean values for all age groups were as follows: adults,  $M = 0.42, SEM = 0.037$ ; 5- to 7-year-olds,  $M = 0.81, SEM = 0.037$ ; 8- to 10-year-olds,  $M = 0.50, SEM = 0.036$ .)

Our analysis of the response latency differences between our main task and the baseline response latency task also revealed main effects of scrambling level [ $F(2, 116) = 98.4, p < 0.001$ , partial  $\eta^2 = 0.63$ ] and age group [ $F(2, 58) = 4.67, p = 0.013$ , partial  $\eta^2 = 0.14$ ]. We examined both of these main effects using post hoc two-tailed paired-samples  $t$  tests implemented in JASP, with Bonferroni-corrected  $p$  values. The main effect of image scrambling was the result of significant differences between all levels of image scrambling, whereas the main effect of age group was the result of significantly slower response latencies from 5- to 7-year-olds relative to adults ( $t = 3.01$ , Cohen's  $d = 0.38, p = 0.012$ ). The interaction between scrambling level and age group did not reach significance,  $F(4, 116) = 1.68, p = 0.158$ , partial  $\eta^2 = 0.055$ . We display the average median response latency across all conditions and age groups in Figure 9.

## Discussion

Using a real/artificial food task, we examined how children's reliance on local and global visual

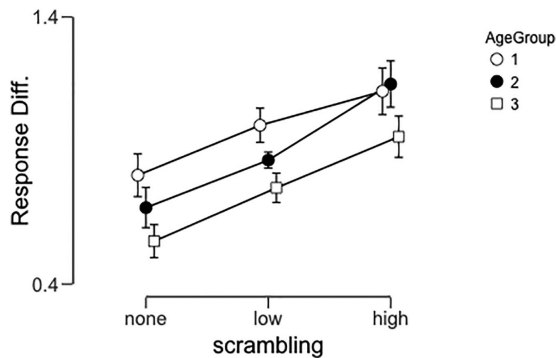


Figure 9. Average response latency as a function of age group (5- to 7-year-olds: Group 1, 8- to 10-year-olds: Group 2, and adults: Group 3) and scrambling level (none, low, and high). Error bars indicate 95% confidence intervals.

information changes across middle childhood. In all three of our tasks, we found that children in all age groups were capable of performing at well above-chance levels, suggesting that this task was a useful means of examining material perception in our target age range. We think accurate performance in this task likely reflects children's expectations regarding the appearance of edible food items and a useful vocabulary for the material appearance of real and artificial food that generally allows them to distinguish between these two categories. Our results across all three experiments demonstrate that the development of children's use of visual information for material perception is not uniform across spatial scales. This is evident in several different features of our data. In terms of accuracy, the critical two-way interaction we observed in [Experiment 1](#) was not evident (at least not as robustly) in either of the following two experiments in which we used image scrambling to disrupt large-scale visual features. One strong interpretation of this outcome is that children do not suffer disproportionately from the removal of small-scale, local features in material perception tasks. We think our data do not unequivocally support this account, however. We observed marginal trends for the same interaction in both [Experiments 2](#) and [3](#), which suggest to us that it may be more likely that the disproportionate cost of image manipulation is larger when fine details are disrupted but that young children may suffer from a smaller disproportionate cost when larger-scale features are disrupted. A direct comparison across these two types of image manipulation with a much larger sample could help us work out if this is the case or not, but the present data do not lend themselves to such an analysis. In particular, we have made no attempt here to match the different levels of blur/scrambling in a meaningful way, so including these in an omnibus ANOVA would be inappropriate. Without this direct comparison, it is difficult to rule out the possibility

that image distortions generally affect younger children more profoundly. Still, we suggest that these results provide useful data supporting the hypothesis that young children rely more heavily on small-scale visual features for material perception than older children and adults do. We note that the various main effects of age that we observed in our response latency data across all three of our experiments are broadly consistent with this account as well. In [Experiment 1](#), we observed significant differences between all three age groups, indicating continued changes in efficiency across middle childhood and into adulthood. In both [Experiments 2](#) and [3](#), however, we only observed significant differences between our youngest age group and adults, suggesting that 8- to 10-year-olds are quantitatively adult-like in terms of their ability to cope with the loss of large-scale visual features. These results do not suggest a disproportionate cost to processing efficiency but rather that there are different developmental trajectories for efficient use of small-scale and large-scale visual features.

In both cases, our data support the hypothesis that children's visual processing follows a local-to-global trajectory during middle childhood. Unlike previous results using schematic stimuli (e.g., Navon figures), we find that children exhibit similar information biases across middle childhood when asked to work with images that are not explicitly hierarchical. Instead, the real/artificial food images we presented to children in these tasks were typical natural images that contained task-relevant visual features across spatial scales. This suggests that the development of global processing during middle childhood is not solely reflected in children's emerging capacity to integrate discrete shape elements into a larger gestalt ([Nayar et al., 2015](#)) but rather applies more generally to the measurement of visual features that extend across large portions of visual space. With regard to material perception specifically, the current results also serve as an important extension of our previous work examining children's ability to use visual summary statistics for material categorization ([Balas, 2017](#)). As we previously noted in the introduction, imposing texture synthesis algorithms like the Portilla-Simoncelli model often involves a commitment to a particular spatial scale, primarily disrupting large-scale visual features more than small-scale features. Although our previous results may thus largely depend on the availability of local visual features, continued examination of how children process material properties specifically and texture more generally would almost certainly benefit from explicit manipulation of the spatial scale parameters intrinsic to the texture models used to manipulate stimuli. A similar approach was adopted by [Freeman and Simoncelli \(2011\)](#) to argue that area V2 encodes and represents local summary statistics, but more generally the technique could be adapted to examine changes in

children's use of information at different spatial scales. Overall, using a range of techniques for controlling the visual features available to children and adults in natural images (including blurring, scrambling, and texture-synthesis approaches) will likely offer more insights into how information use and perceptual strategies change during childhood.

An important issue regarding the overall local-to-global developmental hypothesis is the extent to which this is a general property of visual development or whether this trajectory unfolds at different rates (or perhaps does not happen at all) when we consider other recognition tasks. At the outset of the current study, for example, we were interested in determining the nature of information use for material perception specifically, but do our results instead reflect something more universal or processes other than material perception in particular? For example, one could also imagine that performance in this real/artificial food task depends to some degree on recognizing the food items depicted on each trial. If this were the case, scrambling manipulations could potentially disrupt the recovery of that recognition process, leading to downstream effects on the material judgment that we have focused on. We are not aware of data demonstrating this specific relationship between object recognition and material perception in the context of food items, and in general, material properties are recoverable by children and adults from images that do not clearly depict recognizable objects (Balas et al., 2017). Nonetheless, this is just one example of how a more general local processing bias could underlie the results we have reported here. Our current data do not allow us to speak directly to this point, but the emergence of global processing in other domains provides some evidence that there may not be just one local-to-global shift in visual processing. In particular, children's holistic processing of faces, usually indexed by their performance in a version of the composite face task (Murphy, Gray & Cook, 2017), appears to develop earlier than we would expect based on our data. The emergence of adult-like holistic processing appears to be evident at 4 years of age in some reports (De Heering, Houthuys, & Rossion, 2007; Pellicano, Rhodes, & Peters, 2006), and there is ongoing debate regarding whether there is any meaningful change at all in local versus holistic face processing during childhood (Crookes & McKone, 2009). At the very least, adult-like global face processing at the age of 4 is a far cry from the slower developmental trajectory we observed here. This suggests that material perception is not simply one more by-product of a system-wide change in visual development. An explicit comparison between local and global processing biases for material perception and other texture- or shape-based recognition tasks would be an important extension of our work here, and even examining different types of material perception

in more depth could also yield interesting results. In the current study, we relied on an intuitive judgment regarding food to ensure that children understood our task and also had to successfully interpret a range of different materials. Explicitly examining how children use local and global information to perceive gloss or roughness may help link the use of local and global information to more specific geometric and photometric cues that signal specific material properties.

Overall, our results demonstrate that children's material perception is subject to a fairly slow shift in processing strategy that favors local information early in development. Children's processing of visual texture thus appears to follow a similar trajectory as their ability to integrate discrete visual elements across extended image regions (Kovacs et al., 1999; Nayar et al., 2015). We argue that a key challenge for visual development research is understanding in more depth what global processing *is*. Here, we have discussed our results in terms of integration of visual features across large spatial scales, but the nature of that integration remains largely unknown. Even in adult participants, it is clear that there are integrative processes that make true "metamers" of natural scenes difficult to come by (Wallis, Bethge, & Wichmann, 2016), and examining the emergence of global processing during childhood may be an important means of establishing what visual integration is in more detail. We suggest that material perception is potentially extremely valuable in this regard, and the current study provides useful information about how it develops during childhood with an eye toward probing deeper issues regarding the development of visual recognition.

*Keywords: material perception, visual development, visual recognition*

## Acknowledgments

Supported by NSF grant BCS-1727427 awarded to BB. Special thanks to all the families who volunteered to participate in both experiments, and also to Dan Gu for technical support.

Commercial relationships: none.  
Corresponding author: Benjamin Balas.  
Email: Benjamin.balas@ndsu.edu.  
Address: Department of Psychology, North Dakota State University, Fargo, ND, USA.

## References

- Anderson, B. L., & Kim, J. (2009). Image statistics do not explain the perception of gloss and lightness. *Journal of Vision*, 9(11):10, 1–17, <https://doi.org/10.1167/9.11.10>.

- Arce-Lopera, C., Masuda, T., Kimura, A., Wada, Y., & Okajima, K. (2012). Luminance distribution modifies the perceived freshness of strawberries. *I-Perception*, *3*, 338–355.
- Balas, B. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research*, *46*, 299–309.
- Balas, B. (2017). Children's use of visual summary statistics for material categorization. *Journal of Vision*, *17*, 22.
- Balas, B., Conlin, C., & Shipman, D. (2017). Summary-statistics and material categorization in the visual periphery. *Transactions on Applied Perception*, *14*, 8.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18, <https://doi.org/10.1167/9.12.13>.
- Baumgartner, E., & Gegenfurtner, K. R. (2016). Image statistics and the representation of material properties in the visual cortex. *Frontiers in Psychology*, *7*, 1185.
- Baumgartner, E., Wiebel, C. B., & Gegenfurtner, K. R. (2013). Visual and haptic representations of material properties. *Multisensory Research*, *26*, 429–455.
- Bergmann, W. M., & Kappers, A. M. (2007). Haptic and visual perception of roughness. *Acta Psychologica*, *124*, 177–189.
- Briand, T., Vacher, J., Galerne, B., & Rabin, J. (2014). The Heeger & Bergen pyramid based texture synthesis algorithm. *Image Processing Online*, *4*, 276–299.
- Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition*, *111*, 219–247. doi:10.1016/j.cognition.2009.02.004
- De Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite-face effect. *Journal of Experimental Child Psychology*, *96*, 57–70.
- Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science*, *4*, 233–251.
- Elleberg, D., Hansen, C., & Johnson, A. (2012). The developing visual system is not optimally sensitive to the spatial statistics of natural images. *Vision Research*, *67*, 1–7.
- Fleming, R. W. (2013). Visual perception of materials and their properties. *Vision Research*, *94*, 62–75.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*, 1195–1201.
- Heeger, D. J., & Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. *Proceedings of the 22nd Annual Conference on Computer Graphics & Interactive Techniques*, *30*, 229–238.
- Jacobs, R. H. A. H., Baumgartner, E., & Gegenfurtner, K. R. (2014). The representation of material categories in the brain. *Frontiers in Psychology*, *5*, 146.
- JASP Team. (2018). JASP (Version 0.8.6) [Computer software], <https://jasp-stats.org/faq/how-do-i-cite-jasp/>.
- Kimchi, R., Hadad, B., Behrmann, M., & Palmer, S. (2005). Microgenesis and ontogenesis of perceptual organization. *Psychological Sciences*, *16*, 282–290.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*, 1.
- Kovacs, I., Kozma, P., Feher, A., & Benedek, G. (1999). Late maturation of visual spatial integration in humans. *Proceedings of the National Academy of Sciences*, *96*, 12204–12209.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavioral Research Methods*, *44*, 978–990.
- Meissner, T. W., Prufer, H., Nordt, M., Semmelmann, K., & Weigelt, S. (2018). Development of face detection in preschool children. *International Journal of Behavioral Development*, *42*, 439–444.
- Murphy, J., Gray, K. L. H., & Cook, R. (2017). The composite face illusion. *Psychonomic Bulletin & Review*, *24*, 245–261.
- Nayar, M. S., Franchak, J., Adolph, K., & Kiorpes, L. (2015). From local to global processing: The development of illusory contour perception. *Journal of Experimental Child Psychology*, *131*, 38–55.
- Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science*, *9*, 270–277.
- Poirel, N., Simon, G., Cassotti, M., Leroux, G., Perchev, G., Lanoë, C., . . . Houdé, O. (2011). The shift from local to global visual processing in 6-year-old children is associated with grey matter loss. *PLoS ONE*, *6*(6), e20879.
- Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–71.

- Scherf, K. S., Behrmann, M., Kimchi, R., & Luna, B. (2009). Emergence of global shape processing continues through adolescence. *Child Development, 80*, 162–177.
- Schmidt, F., & Fleming, R. W. (2018). Identifying shape transformations from photographs of real objects. *PLoS ONE, 13*(8), e0202115.
- Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). Material perception: What can you see in a brief glance? *Journal of Vision, 9*, 784.
- Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-world images. *Journal of Vision, 14*, 1–24.
- Sireteanu, R., & Rieth, C. (1992). Texture segregation in infants and children. *Behavioural Brain Research, 49*, 133–139.
- Stojanowski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic scrambling. *Journal of Vision, 14*, 12, 1–16.
- Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology, 6*, 632–650.
- Wallis, T. S. A., Bethge, M., & Wichmann, F. A. (2016). Testing models of peripheral encoding using metamers in an oddity paradigm. *Journal of Vision, 16*(2), 4.
- Wiebel, C. B., Valsecchi, M., & Gegenfurtner, K. R. (2014). Early differential processing of material images: Evidence from ERP classification. *Journal of Vision, 14*(7):10, 1–13.