



PAPER

Ensemble perception of size in 4–5-year-old children

Timothy D. Sweeny,¹ Nicole Wurnitsch,² Alison Gopnik² and David Whitney^{2,3}

1. Department of Psychology, University of Denver, USA

2. Department of Psychology, University of California – Berkeley, USA

3. Vision Science Group, University of California – Berkeley, USA

Abstract

Groups of objects are nearly everywhere we look. Adults can perceive and understand the ‘gist’ of multiple objects at once, engaging ensemble-coding mechanisms that summarize a group’s overall appearance. Are these group-perception mechanisms in place early in childhood? Here, we provide the first evidence that 4–5-year-old children use ensemble coding to perceive the average size of a group of objects. Children viewed a pair of trees, with each containing a group of differently sized oranges. We found that, in order to determine which tree had the larger oranges overall, children integrated the sizes of multiple oranges into ensemble representations. This pooling occurred rapidly, and it occurred despite conflicting information from numerosity, continuous extent, density, and contrast. An ideal observer analysis showed that although children’s integration mechanisms are sensitive, they are not yet as efficient as adults’. Overall, our results provide a new insight into the way children see and understand the environment, and they illustrate the fundamental nature of ensemble coding in visual perception.

Research highlights

- Specialized ensemble-coding mechanisms are in place at 4–5 years of age.
- Sensitive group perception can occur in spite of poor sensitivity to individuals.
- Despite high sensitivity, ensemble coding is still not fully developed in young children.
- Children may use ensemble percepts to grasp the gist of a scene.

Introduction

Take a trip to the market and you will appreciate how readily people perceive and evaluate groups. Dozens of oranges are seen not as individual objects, but as collectives of small and large, groups of clementines and Valencia. Gestalt impressions such as these provide a window into the phenomenology of visual experience, revealing that much of awareness is at the level of the

group (Koffka, 1935). But there is an apparent paradox about this kind of perception. The capacity limits of attention and memory only allow us to perceive and remember precise information about a few objects at a time (Awh, Barton & Vogel, 2007; Luck & Vogel, 1997; Scholl & Pylyshyn, 1999; Simons & Levin, 1997; Whitney & Levi, 2011). How then, does the visual system overcome these bottlenecks and enable us to see what a group of objects is like all at once, as a collective?

To see groups, adults engage a process known as ensemble coding. In ensemble coding, information about multiple objects is compressed into a summary statistic – a singular representation that provides access to the group in the form of a gist percept (for reviews, see Alvarez, 2011; Whitney, Haberman & Sweeny, 2014). This compression is what allows ensemble codes to overcome capacity limits, providing more information about groups than would be possible by inspecting each member of a set, one after another. In fact, with ensemble coding, precise information about individuals,

Address for correspondence: Timothy D. Sweeny, University of Denver, Department of Psychology, Frontier Hall, 2155 South Race Street, Denver, CO 80208, USA; e-mail: timothy.sweeny@du.edu

like each orange's unique size, is lost in favor of the group percept, like the average size of all the oranges in the group (Ariely, 2001; Haberman & Whitney, 2009, 2011). Moreover, ensemble codes need not be drawn from every member in a group (Dakin, Bex, Cass & Watt, 2009) – rapid averaging across a subset of members is enough to provide surprisingly high sensitivity, even when there is insufficient time to serially attend to individual group members (e.g. Sweeny, Haroz & Whitney, 2013). In fact, this averaging process overcomes noise in local representations (Alvarez, 2011), allowing adults to perceive crowds more precisely than they perceive individuals (Sweeny, Haroz, & Whitney, 2013). Efficient and sensitive, ensemble codes underlie an important part of conscious phenomenology, providing the gist experience that allows adults to regard sets of objects as holistic groups with particular collective characteristics. We might think of ensemble perception as involving a kind of 'lantern consciousness' in which many objects are seen at once, rather than the more focused 'spotlight consciousness' that comes with attention to individual objects (Gopnik, 2009).

Less is known about what visual experience is like early in life. Do children see their surroundings like adults, appreciating multiple objects as groups with collective properties? Or is children's visual experience simply noisier, or less cohesive in general than adults'? Children's perception is certainly coarse – spatial and temporal resolution develop slowly (Benedek, Benedek, Kéri & Janáky, 2003; Ellemberg, Lewis, Liu & Maurer, 1999; Farzin, Rivera & Whitney, 2010, 2011; Jeon, Hamid, Maurer & Lewis, 2010), and crowding, the impairment in perceiving an individual item when it is surrounded by clutter, is particularly strong in childhood (Jeon *et al.*, 2010; Semenov, Chernova & Bondarko, 2000). Moreover, children's attentional capacities are also limited compared to those of adults, particularly their capacities for focused, endogenous top-down attention (Enns & Girgus, 1985; Goldberg, Maurer & Lewis, 2001; Johnson, 2002; Ristic & Kingstone, 2009). These limitations make it easy to imagine that, when children are viewing many objects at once, their visual experience would be noisy and fractured rather than holistically coherent.

On the other hand, studies of cognitive development show that even infants are remarkably sensitive to statistical and relational features of scenes, and that they seem to discriminate scenes based on judgments about those features (e.g. Spelke, Breinlinger, Macomber & Jacobson, 1992; Fiser & Aslin, 2002; Kirkham, Slemmer & Johnson, 2002; Xu & Garcia, 2008). Furthermore, children are adept at making inferences

from subtle statistical cues even by preschool age (see Gopnik & Wellman, 2012, for a review).

Children might, in fact, be able to mitigate their perceptual and attentional limitations by engaging ensemble-coding strategies. By pooling independent samples of similar objects, children may be able to average out the noise, improving their visual sensitivity to perceive groups of objects in scenes. Here, we sought to test these competing hypotheses by determining whether or not ensemble-perception mechanisms are in place in childhood.

Specifically, we asked: when 4–5-year-old children see many objects at once, do they determine the average object size in the group by considering information from multiple objects (i.e. ensemble coding), or do they make coarse judgments about the group based only on a single object? Judgments of ensemble size based on just one object are likely to be coarse, and not ideal, because the size of the sampled object may not necessarily be representative of the entire group. We investigated perception of size for two reasons. First, ensemble-size perception is robust and well understood with adults (e.g. Ariely, 2001; Chong & Treisman, 2003). Second, size is a relatively simple visual feature in the hierarchy of visual processing (Desimone & Schein, 1987; Dumoulin & Hess, 2007). We therefore reasoned that if *any* ensemble coding is operative in childhood, it is likely to be manifested in a judgment of size.

While the perception of basic collective properties of homogeneous groups, like numerosity, comes online early in life (Zosh, Halberda & Feigenson, 2011), it is still unclear whether children possess the ensemble processes necessary to summarize properties of heterogeneous groups. This distinction is critical. Only ensemble coding affords humans the ability to *average* information across many individuals that differ in appearance, telling us not just how much stuff there is, but what that stuff is like. Thus, the current research represents an important advance in understanding the development of complex group perception.

Why would it matter if ensemble mechanisms were or were not in place in childhood? Adults are already known to use ensemble coding to overcome their own perceptual limitations. Specifically, pooling information into gist percepts allows adults to recover information about individuals that would otherwise be inaccessible or unrecognizable due to crowding (Fischer & Whitney, 2011), visual short-term memory capacity (Alvarez & Oliva, 2008; Haberman & Whitney, 2011), attentional capacity (Chong & Treisman, 2003), or even congenital deficits in recognition ability (Yamanashi Leib, Puri, Fischer, Bentin, Whitney & Robertson, 2012). If

children also possess ensemble mechanisms, they could, in principle, use them to overcome their own capacity limits. This would be especially useful at 4–5 years of age, when children’s perception of individual objects is already more limited than that of adults’, and is constrained by gradually developing spatial resolution (Farzin *et al.*, 2010), selective attention (Enns & Girgus, 1985; Goldberg *et al.*, 2001; Johnson, 2002; Ristic & Kingstone, 2009), and visual working memory capacity (Cowan, AuBuchon, Gilchrist, Ricker & Sauls, 2011; Simmering, 2012; Wilson, Scott & Power, 1987). For these reasons, we tested 4–5-year-old children. Testing children at this age has an additional benefit – we could ask them to make explicit judgments of groups in the same experimental paradigms as those used with adults, which would allow us to directly compare the strength of children’s and adults’ ensemble integration capabilities.

If ensemble perception mechanisms are in place early in life, then just like adults, children should be able to use ensemble coding to extract information about average size in a group of objects. Alternatively, finding that children are unable to access summary information at the level of the group would reveal that their visual experience is even more limited than previously thought. As such, our results should have far-reaching implications, not just for characterizing the development of ensemble coding for the first time, but also for providing new insights into the phenomenology of visual experience in childhood.

Experiment 1

Methods

Observers

Ten children ($M = 4.67$ years, $SD = 0.58$ years, five females and five males), and 10 adults ($M = 28.9$ years, $SD = 4.6$ years, five females and five males) participated in the experiment.

Stimuli

Children were seated in front of a laptop but were free to move during the experiment, so the visual angles noted below represent the size of the stimuli from an average distance of 57 cm. These values would have varied between -15% and $+20\%$ given 10 cm of movement forward or backward. Because children are more likely to cooperate when an experiment is presented as a game, we created a cartoon scene which we described as a chance to help a ‘hungry monkey’ determine which of two trees had the largest oranges (Figure 1a). Every aspect of this cartoon scene was prepared in Photoshop (Adobe Photoshop CS5 Version 12.0). Oranges had an orange hue. The two trees were identical ($16.1^\circ \times 19.2^\circ$), each with a green top and a brown trunk. The edges of the trees abutted one another. The cartoon monkey ($2.1^\circ \times 4.81^\circ$) had brown fur. Green grass lined the bottom border of the screen ($30.1^\circ \times 2.61^\circ$). Oranges were presented along an

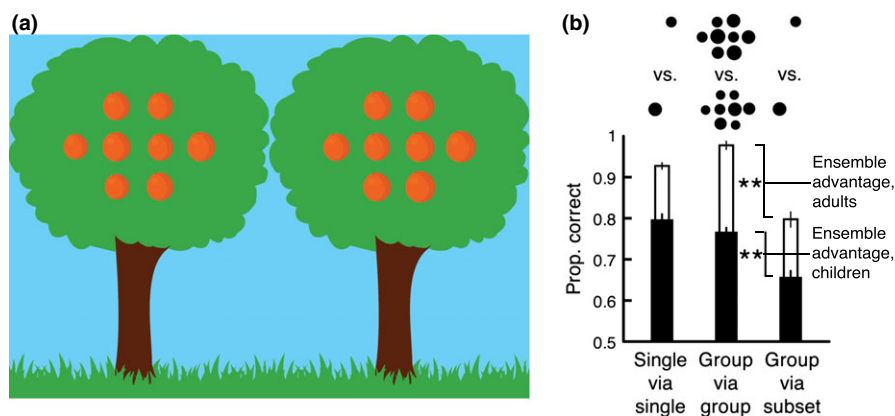


Figure 1 (a) Each tree had a range of differently sized oranges. In the group condition, multiple oranges were available within each tree for determining which had the largest oranges, overall. In the single condition (not shown), only a randomly selected pair of oranges from different sets of full groups was visible. Both the single-via-single analysis and the group-via-subset analysis were conducted on the same data from the single condition. (b) Children (black bars) selected the tree with the largest oranges, overall, more often when the full sets of oranges were visible (shown in the group-via-group analysis) compared to when only a single orange (a subset) was visible in each tree (shown in the group-via-subset analysis). This occurred despite high sensitivity for discriminating individual oranges observed in the single-via-single analysis. Adults (white bars) showed the same pattern. ** represents $p < .01$. Error bars reflect ± 1 SEM (adjusted for within-observer comparisons).

invisible grid ($9.85^\circ \times 6.6^\circ$) centered in each tree. The entire scene appeared over a blue background.

On each trial, we generated groups of eight oranges for each tree. Each tree contained four different orange sizes, and one tree (randomly determined as the right or left) had slightly larger oranges, on average. Individual orange sizes ranged from 1.69° to 2.58° across trials. On each trial, the range of orange sizes within each tree (0.294°) and the difference in average orange size between the trees (0.147°) remained constant. Each orange in the left tree was always a different size from any orange in the right tree. Across trials, the average orange size collapsed across both trees ranged from small to large (1.91° , 1.97° , 2.02° , 2.06° , 2.11° , 2.16° , 2.21° , 2.26° , 2.31° , or 2.36°). On a given trial, the average orange size in either tree was always $.074^\circ$ smaller or larger than one of these values. Oranges based around these 10 sizes were each generated four times for a total of 40 trials across two blocks of 20 trials each. The variability in orange sizes within each tree and the overlapping sizes between the trees served an important purpose – it ensured that sampling multiple oranges from each tree (i.e. building ensemble representations) would lead to better judgments of average size than simply comparing a single orange from each tree (see Table 1 for a detailed description of orange sizes, conditions, and analyses in two hypothetical trials).

Procedure

Each trial began with a cartoon monkey standing equidistant from the pair of trees. Each observer was free to look directly at the oranges for an unlimited duration, and

then indicated which tree had the larger oranges, overall, by moving the monkey (4.31° leftward or rightward) under one of the trees using the left or right arrows on a keypad. Next, the observer pressed the spacebar once to finalize his or her response and again to start a new trial. A practice block of four trials preceded the main experiment. In this practice block, the experimenter explained (1) that the monkey was hungry and preferred to eat large oranges, (2) that the observer could help the monkey by moving him to the tree with the larger oranges, overall, and (3) that the observer should consider the sizes of all of the oranges visible to him or her when making a decision.

On half of the trials, full groups of eight oranges appeared in each tree (Figure 1a) – the *group condition*. In our analysis of this condition, we calculated the proportion of trials in which each observer selected the tree with the largest average orange size, presumably via an evaluation of the groups. We refer to this as the *group-via-group analysis*.

However, the main objective of this investigation was not simply to determine whether children performed well when comparing groups. Rather, it was to determine *how much* information they used to evaluate these groups. To answer this question, we included a control condition – the *single condition* – to determine what performance would have looked like if children judged the groups simply by comparing a single orange from each tree. Specifically, on the other half of trials, we again generated full groups of eight oranges for each tree, but only displayed a single randomly selected orange (i.e. a subset) from each tree. We recorded the average orange size in each full group as well as the sizes of the single

Table 1 Conditions, analyses, individual oranges sizes, and group average sizes from two hypothetical trials in Experiment 1

Condition	Analysis	Tree	Orange sizes (1–8)								Group Avg
Group	Group -via- group	L	68	68	<i>72</i>	<i>72</i>	<i>76</i>	<i>76</i>	82	82	74
		R	74	74	<i>78</i>	<i>78</i>	82	82	<i>86</i>	<i>86</i>	80
Single	Single -via- Single Subset	L	<i>86</i>	86	<i>90</i>	<i>90</i>	<i>94</i>	<i>94</i>	<i>98</i>	<i>98</i>	92
		R	<i>80</i>	<i>80</i>	<i>84</i>	<i>84</i>	<i>88</i>	88	<i>92</i>	<i>92</i>	86

Note: In each trial of each condition, a separate group of eight oranges was generated for the left and right tree. For simplicity in this table, the sizes (diameter, in pixels) of the oranges in each tree are listed in order from smallest to largest. In the group condition, in which all eight oranges were presented and visible in each tree, we analyzed how well children and adults evaluated the average size in each tree – the group-via-group analysis. In this trial of the group condition, the right tree happens to have the larger average orange size (80 pixels). Note, however, that the largest orange in the left tree was larger than the smallest orange in the right tree. This illustrates how comparing a random pair of oranges in each tree would not always lead to a correct response about the groups. In the single condition, eight oranges were again generated for each tree, but only one randomly selected orange (in bold) was visible in each tree. The rest of the oranges (in italics) were not presented. We conducted two analyses – the single-via-single analysis and the group-via-subset analysis – on the same data from the single condition. The single-via-single analysis allowed us to evaluate sensitivity for comparing individual oranges. Here, the orange in the right tree (92 pixels) is the largest, and selecting the right tree leads to a correct answer in the single-via-single analysis. The group-via-subset analysis allowed us to determine what performance in the group condition would have been like had observers based their estimates on a single randomly selected orange in each tree. Here, accurately comparing this randomly selected pair of oranges would not have led to a correct response about the groups. Better performance in the group condition relative to this simulated performance in the group-via-subset analysis would mean that observers' did not simply compare random pairs of oranges in the group condition.

randomly displayed oranges. Trials from both conditions were intermixed within both blocks.

We analyzed the data from this same set of trials in the single condition in two different ways. In the *group-via-subset analysis*, we calculated the proportion of trials in which each observer's selection (made via the subsets, each of which always contained just one orange) happened to also be the tree with the largest group average. This calculation simulated what performance would have looked like in the group condition if observers had based their judgments on a single orange in each tree. The purpose of this single condition, as analyzed in this way, was not to ask observers to make judgments of groups that they could not see, which would be impossible. Rather, this analysis is analogous to an empirical simulation of what performance in the group-via-group condition would have looked like if observers had not engaged crowd perception, combining information from several objects, but instead had based their judgment on a single object that we randomly selected for them.

In the *single-via-single analysis*, we calculated the proportion of trials in which each observer selected the tree that contained the subset with the largest orange. That is, within the single condition, we simply determined how often the observer chose the tree with the larger single orange. This latter calculation allowed us to measure baseline sensitivity to individual size differences (see Table 1 for a detailed description of these analyses in two hypothetical trials).

If observers utilize ensemble coding, they will select the tree with the largest oranges, overall, more often when the full groups are visible. Performance should be better in the group condition than in the group-via-subset analysis of the single condition – the *ensemble advantage*. This should occur despite high baseline sensitivity for comparing individual oranges and concomitant high performance in the single-via-single analysis.

Results

Children showed high sensitivity for comparing the sizes of individual oranges; they were very accurate at determining which single orange was larger (the single-via-single analysis; Figure 1b). They were also highly precise and accurate when they compared groups of oranges (the group-via-group analysis). Performance in the single-via-single analysis and in the group condition were strongly correlated ($R^2 = .868$, $p < .01$). Most importantly, children's comparisons of groups would have been less precise if they had been made via a single random pair of oranges; children selected the tree with the larger oranges less often when only a single randomly selected orange was visible from each group ($t[9] = 4.13$,

$p < .01$, Figure 1b). This difference – the *ensemble advantage* – directly shows that when children made judgments about groups of oranges, they did so by evaluating information from multiple oranges and not just by using a single orange within each group.

Adults were excellent at discriminating the sizes of individual oranges, making even fewer errors than children in the single-via-single analysis ($t(18) = 2.61$, $p < .05$). Adults were also adept at perceiving groups, and they made more precise comparisons of groups via consideration of multiple oranges in each tree (the *ensemble advantage*: $t[9] = 4.91$, $p < .01$, Figure 1b). Overall, the similar pattern of results from children and adults suggests that ensemble coding is in place even at 4–5 years of age. Thus, while children's perception is certainly noisier, consistent with previous results (Benedek *et al.*, 2003; Ellemberg *et al.*, 1999; Farzin *et al.*, 2010, 2011; Jeon *et al.*, 2010), they appear to employ an ensemble-coding strategy which averages out noise and achieves a more precise percept of the group than would otherwise be possible.

Exactly how refined are children's group perception mechanisms? In order to properly evaluate children's sensitivity to groups, it was first necessary to determine whether their lower *overall* performance was due to perceptual limitations, or instead due to lapses in attention or motivation. We accomplished this by deconstructing data from the single-via-single analysis, evaluating performance across the wide range of size differences that resulted from randomly selecting and displaying a single orange from each tree. Specifically, we separately concatenated data from children and adults, and then measured each group's discrimination performance as a function of the differences in size between the two randomly visible oranges. For both children and adults, the proportion of correct responses increased with larger differences in size, and these increases were well fit by logistic functions (children; $R^2 = .852$, adults; $R^2 = .911$, Figure 2a). We then used these logistic fits to separately derive probability density functions representing children's and adults' sensitivities to size differences. Specifically, using our logistic fits from Figure 2a, we simply determined the increase in percent correct that accompanied incremental increases (in 0.2 degree steps) in the size difference of the two oranges being compared. This yielded half of a Gaussian-shaped distribution, which we then mirrored to produce a full probability density function separately for children and adults. The standard deviations of these density functions were remarkably similar (children; $SD = .098^\circ$, adults; $SD = .10^\circ$).

This similarity has two important implications. First, it indicates that children's sensitivity for comparing the sizes of individual objects is fully developed at 4–5 years

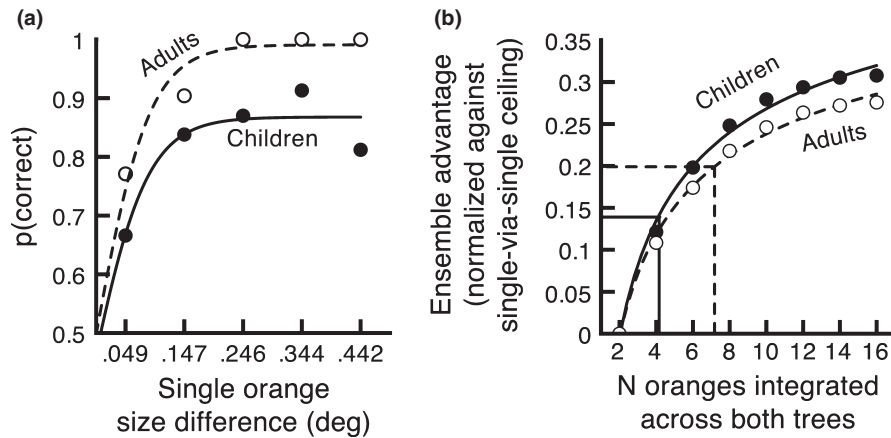


Figure 2 (a) Children's and adults' abilities to discriminate size in the single-via-single analysis. Both groups' performances improved with larger differences in size and were well fit by logistic functions. We used these logistic fits to simulate early-stage noise in our ideal observer analysis. (b) Number of oranges that children and adults integrated across both trees, estimated from our ideal observer analysis. We used a linear pooling algorithm to predict the ensemble advantages that children (filled circles) and adults (open circles) would gain from integrating different numbers of oranges in each tree. The intersection of children's and adults' observed ensemble advantages (horizontal lines) with power functions fitted to our predicted ensemble advantages provided estimates of the number of oranges each group integrated across both trees (vertical lines).

of age. Second, it shows that children's lower baseline level of performance reflected non-perceptual limitations, meaning that any comparison of children's and adults' sensitivity to groups should be made independently of this baseline difference. That is, the differences between children and adults were not due to basic differences in their perceptual abilities, but were more likely due to lapsing, inattentiveness, or a decision stage process. Children might, for example, be less good than adults at actually implementing the decision to pick a particular tree.

Ideal observer analysis

How sensitive are children's group perception mechanisms compared to adults'? How many oranges did children and adults integrate into their ensemble codes? To answer these questions, it was first necessary to calculate children's and adult's improvements in proportion correct – their *ensemble advantages* – independently of baseline differences in performance. Specifically, we divided each observer's ensemble advantage by his or her ceiling performance in the single-via-single analysis. We then performed an ideal observer analysis to estimate the number of oranges children and adults would have needed to integrate to produce these observed ensemble advantages. This analysis is straightforward and conservative; similar linear pooling models have successfully characterized ensemble coding in adults (Parkes, Lund, Angelucci, Solomon & Morgan, 2001; Sweeny, Haroz & Whitney, 2012). This

assumption of linearity is reasonable considering (1) the small range of sizes in this experiment (Chong & Treisman, 2003), and (2) the very similar results we obtained when we ran this same ideal observer analysis using values adjusted to reflect the compressive-non-linear nature of size perception (Teghtsoonian, 1965).

We began our ideal observer analysis by simulating performance in the group-via-group analysis and the group-via-subset analysis. First, we generated eight orange sizes for each tree using actual values from the experiment. Next, we simulated noisy perception of each orange by (1) centering Gaussian-shaped sampling distributions at each simulated orange size, each with a standard deviation equal to our estimates of children's or adults' single-orange sensitivity (see above) and (2) sampling a random value from each distribution. This sampling produced groups of eight noise-perturbed oranges in each tree. Last, we simulated performance that would have occurred with different numbers of oranges integrated. We started by randomly selecting subsets (1–7) of our simulated oranges or the full group of eight oranges from each tree. For each simulated subset in each tree, we calculated the average orange size (the linear mean of the subset) and selected the tree with the largest average value as the simulated response. We did not add noise in this integration stage. As in our experiment, we noted whether the simulated response (based on each subset) matched the correct response that would be obtained from comparing the averages of the full groups in each tree.

We completed this procedure 50,000 times and determined the proportion of these trials in which each

number of oranges integrated (e.g. four across both trees, five across both trees, etc.) produced the correct response. We then calculated the difference between these values and the proportion of correct responses that would have been obtained from using a subset of one orange from each tree. In this way, we simulated the ensemble advantages children and adults would have produced by integrating different numbers of oranges (Figure 2b). We then fit the increase in these simulated ensemble advantages with a power function. The fits were excellent ($R^2 > .977$, $p < .01$ for children and adults). By locating the points where children's and adults' observed ensemble advantages intersected with these fits, we were able to estimate the number of oranges that each group integrated into their respective ensemble codes (Figure 2b). We found that, across both trees, children integrated at least 4.24 oranges and adults integrated at least 7.18 oranges. We obtained similar patterns of results by comparing subsets with different sizes but equivalent totals (e.g. averaging four and two oranges from each tree instead of three and three from each tree). However, for each combination with unbalanced subsets, it would have been necessary to integrate more oranges than with balanced subsets, with even the closest estimates requiring an extra orange to be integrated. While we cannot be sure what approach observers used, especially on any given trial, our approach of using balanced subsets at least provides the most conservative estimate of the number of oranges integrated.

This analysis takes into account the differences in attention, motivation, decision-making and other sources of noise that are reflected in the differences between adults and children on the single-orange task. However, like other ideal-observer simulations of ensemble coding (Myczek & Simons, 2008), ours does not include late-stage noise in the integration process. Accounting for this attribute of real, imperfect observers would only increase estimates of integration efficiency since more samples would need to be pooled to reach the observed enhancement. As such, our simulation provides a conservative, lower-bound estimate of integration efficiency. We also do not know whether there might be developmental differences in this late-stage noise, or whether such differences would influence the relative size of the ensemble advantage for adults and children.

Experiment 2

The children in our first experiment showed a pattern of results consistent with integrating multiple oranges to perceive average size. Nevertheless, their sensitivity to

single oranges was positively correlated with their sensitivity to groups. In order to be sure that children possess a distinct mechanism for evaluating crowds, it is important to demonstrate that sensitivity to crowds persists even when it is dissociated from sensitivity to individual objects. We thus conducted Experiment 2 to determine whether children can be good at perceiving crowds despite poor sensitivity to individuals. Such a result would clearly show that children do indeed possess unique ensemble size representation. In addition, we wanted to determine whether or not sensitive crowd perception persists when children have insufficient time to serially attend to individual oranges and cognitively compute the average size. This alternative strategy could have been possible with the unlimited viewing duration in Experiment 1, and might not qualify as ensemble coding (Myczek & Simons, 2008), in which the average percept is achieved through rapid integration in parallel, bypassing the need for focused attention. Rejection of this kind of focused-attention strategy has been important for providing rigorous evidence of summary statistical perception (Ariely, 2008; Chong, Joo, Emmanouil & Treisman, 2008; Myczek & Simons, 2008), and it is especially relevant here, since larger objects are known to draw bottom-up attention over smaller objects when in groups (Proulx, 2010; Treisman & Gormican, 1988).

To evaluate both of these questions, in Experiment 2, we tested whether the integration we observed in Experiment 1 could occur when children viewed groups for only 1500 msec. We reasoned that, with this shorter viewing time, children would be more likely spread their attention to access the gist of the crowds independent from, and at the expense of, sensitivity to individual oranges. Although 1500 msec may seem relatively long compared to durations typically used in research with adults, it was the shortest duration that children would endure in a round of pilot testing. Importantly, we began Experiment 2 with a preliminary investigation (Experiment 2A), which confirmed that 1500 msec was an insufficient amount of time for children to engage a serial search strategy for the largest orange.

Experiment 2A

Methods

Observers

A new group of 20 children ($M = 4.26$ years, $SD = .877$ years, seven females and 13 males), and a new group of 20 adults ($M = 20.28$ years, $SD = 2.39$ years, 12 females and eight males) participated in the experiment.

Stimuli and procedure

The stimuli and procedure were identical to those in Experiment 1, with oranges displayed until each observer responded. The only exception was that observers were instructed to point to the largest orange on the screen as quickly as possible. The experimenter made sure that each child was looking at the screen before pressing the spacebar to initiate the trial. The experimenter pressed the right- or left-arrow key as soon as the child pointed to the screen, taking care to respond as soon as the child's finger clearly indicated the orange of his or her choice. The experimenter repeated the instructions to find the largest orange and point to it as quickly as possible throughout the experiment. The same procedure was used for testing adults so that any experimenter-imposed response delay would be equivalent across reaction times for children and adults. Children completed 10 trials from the group condition and 10 trials from the single condition, all in a single block of testing. Adults completed twice the number of trials intermixed across two blocks of testing.

Results

To obtain a conservative estimate of reaction times, we removed trials in which RTs were greater than 2.5 standard deviations from each observer's mean reaction time. We conducted this trimming separately for trials where an observer viewed groups and trials where that

same observer viewed a single orange in each tree. We did not exclude trials based on whether or not the observer actually pointed to the largest orange.

Not surprisingly, children responded faster on trials where a single orange was presented in each tree ($M = 2607$ msec, $SD = 780$ msec) compared to when a group was presented in each tree ($M = 2984$ msec, $SD = 1042$ msec), $t(19) = 2.77$, $p < .01$, $d = .619$. Critically, mean RTs on trials where children viewed groups were longer than 1500 msec, $t(19) = 6.37$, $p < .01$, $d = 1.42$. A histogram combining raw RTs from all children clearly shows that, on trials from the group condition, 99.48% of individual RTs were greater than 1500 msec (Figure 3a). There was no relationship between RT in the group condition and age, $R^2 = .087$, *ns*. While some portion of children's reaction times reflects the additional time it took for the experimenter to record his or her response, this added time would have been equivalent for each observer in each condition, and it is reasonable to expect that it would have been relatively small, as the experimenter recorded the child's response as soon as his or her finger touched the screen.

Like children, adults responded faster on trials where a single orange was presented in each tree ($M = 1180$ msec, $SD = 112$ msec) compared to when a group was presented in each tree ($M = 1387$ msec, $SD = 171$ msec), $t(19) = 7.29$, $p < .01$, $d = 1.63$. Unlike with children, adult's mean RTs on trials from the group condition were *shorter* than 1500 msec, $t(19) = 2.81$, $p < .05$,

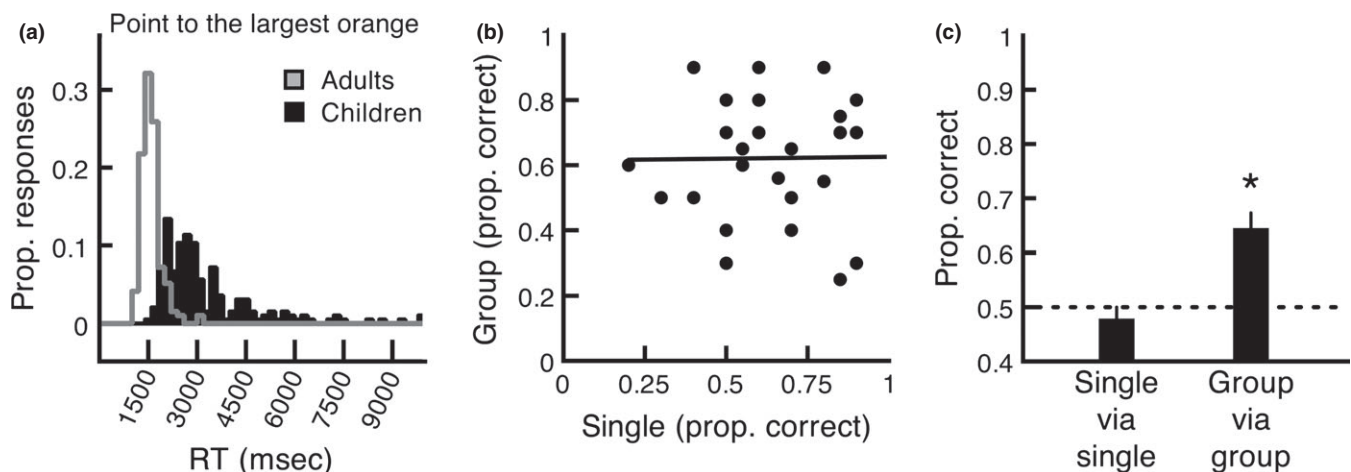


Figure 3 (a) Histograms of children's (black bars) and adults' (gray line) reactions times for pointing to the largest orange on the screen in Experiment 2A. RTs from individual trials were pooled for each group after extreme RTs were excluded separately for each observer (see text for details). Note that children's RTs were overwhelmingly longer than the 1500 msec they had to view the oranges in Experiment 2B. (b) With the shorter presentation time in Experiment 2B, there was no relationship between children's sensitivity to single orange size differences and group size differences. (c) Based on a median split, the children who had the lowest sensitivity for comparing individual orange sizes (not above chance-level performance, left bar) showed higher sensitivity for comparing the sizes of groups. * represents $p < .05$. Error bars reflect ± 1 SEM (adjusted for within-observer comparisons).

$d = .629$. A histogram combining raw RTs from all adults illustrates that 77.2% of individual RTs were less than 1500 msec (Figure 3a). Note that adult RTs included the same experimenter-produced additional delay included in child RTs.

We mention reaction times from adults merely to illustrate the relatively long amount of time it takes children to complete this particular type of visual search. While it would be odd to describe 1500 msec as a brief duration in an experiment designed for adults, this time is indeed brief and challenging for 4–5-year-old children. Most importantly, these results confirm that 1500 msec is an insufficient amount of time for 4–5-year-old children to point to what they believe to be the largest orange on the screen. We can thus be confident that, should children display any sensitivity for perceiving groups displayed for this duration in Experiment 2B, this sensitivity will not have emerged from engaging a serial search strategy, and it should not be related to sensitivity for perceiving single oranges.

Experiment 2B

Methods

Observers

A new group of 27 children ($M = 4.35$ years, $SD = .81$ years, 15 females and 12 males), and 20 adults (the same observers from Experiment 2A) participated in the experiment.

Stimuli and procedure

The stimuli and procedure were identical to those in Experiment 1 in which children were instructed to determine which tree had larger oranges overall, except that the oranges were shown for only 1500 msec. On each trial, the experimenter made sure that each child was looking at the screen before pressing the spacebar to initiate the trial.

Results

Even with the shorter duration, children were above chance for comparing the sizes of individual oranges in the single condition ($M = .642$, $SD = .193$), $t(26) = 3.84$, $p < .01$, $d = .741$, and for comparing groups of oranges in the group condition ($M = .62$, $SD = .183$), $t(26) = 3.42$, $p < .01$, $d = .658$. Unlike with an unlimited presentation time, sensitivity for discriminating mean size in briefly

presented crowds was not related to sensitivity for discriminating the sizes of individual oranges ($R^2 = .0001$, ns , Figure 3b). These results show that children are able to make comparisons of average size in groups without relying on their sensitivity to individual oranges.

The fact that perception of crowds and individuals were independent in this experiment allowed us to more conclusively evaluate the possibility that some observers, as a group, might have been better at extracting information from groups than from single oranges. We did this by dividing the children into two groups using a median split based on single-orange sensitivity. The 13 children who discriminated single oranges with the lowest sensitivity ($M = 47.7\%$ correct, not significantly different from chance-level performance, $t[12] = 0.68$, ns) nevertheless discriminated the average sizes of groups of oranges above chance level ($t[12] = 2.78$, $p < .05$, Figure 3c). Moreover, and surprisingly, these children's group discriminations were actually *better* than their discriminations of a single orange from each tree ($t[12] = 3.13$, $p < .01$, Figure 3c). Overall, these surprising patterns of results show that rapid-acting ensemble perception mechanisms are in place by 4–5 years of age and that they operate independently of sensitivity for perceiving individual objects.

Experiment 3

The findings from Experiments 1 and 2 converge to suggest that children do indeed integrate and summarize size information from multiple oranges to make evaluations of groups – *ensemble coding*. However, it is still possible that they could have used other information related to average size, like increased density or the amount of orange color on the screen, to make their choices. We conducted Experiment 3 to directly rule out these image-based explanations.

Specifically, we manipulated the numerosity of the groups such that one tree always had eight oranges (as in the previous experiments) and the other always had six oranges, with reduced density and total orange color displayed. This design allowed us to conclusively determine whether or not ensemble size integration is available at 4–5 years of age, and if so, how it interacts with orthogonal image-based information.

Methods

Observers

A new group of 20 children ($M = 4.34$ years, $SD = .63$ years, 13 females and seven males), and 20 adults (the

same adult observers from Experiment 2) participated in the experiment.

Stimuli and procedure

Most stimuli and procedures were identical to those in Experiment 1, with oranges displayed until each observer responded. The only exceptions were that (1) we only included the group condition – there were no trials where a single randomly selected orange was visible in each tree, and that (2) only six oranges were visible in one of the two trees, randomly determined to be the left or right tree.

As in the single condition in Experiment 1, we generated a full set of eight oranges for each tree even when only six were displayed in one of them. For the six-orange tree, we randomly selected six oranges from this full set and displayed them in six randomly selected locations out of the eight locations possible. Thus, across trials, the oranges on the tree with six oranges were less densely spaced than those on the tree with eight oranges. Children completed 20 trials from the group condition in a single block of testing. Adults completed 40 trials across two blocks of testing.

Results

When the tree with the largest oranges also happened to be the tree showing the most (eight) oranges, both children and adults easily discriminated average orange size across the groups, and were significantly above chance (children; $t[19] = 5.32$, $p < .01$, $d = 1.18$, adults; $t[19] = 46.9$, $p < .01$, $d = 10.49$, Figure 4). More importantly, when the tree with the largest oranges happened to be the tree with lower numerosity, showing just six oranges, children and adults still discriminated the groups above chance (children; $t[19] = 2.64$, $p < .05$, $d = .59$, adults; $t[19] = 8.50$, $p < .01$, $d = 1.90$, Figure 4). Average size perception was not completely independent of numerosity, density, or the amount of orange on the screen; both children and adults showed reduced sensitivity for comparing average size when the largest-sized oranges were in the less numerous group (children; $t[19] = 3.24$, $p < .01$, $d = .726$, adults; $t[19] = 3.14$, $p < .01$, $d = .702$, compare left bars with right bars in Figure 4).

The important point here, however, is that children can still reliably extract ensemble size information even in the face of strong interference from orthogonal image-based information. Moreover, there were actually no numerosity differences in Experiment 1, and though the density differences in Experiment 3 were larger than those in Experiment 1, the children still showed ensemble perception. Hence, the ensemble advantage in

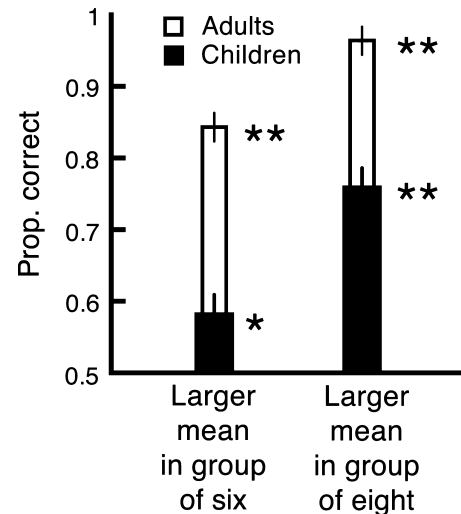


Figure 4 Despite baseline differences in sensitivity, children (black bars) and adults (white bars) were able to indicate which group had the largest oranges, on average, even when that group had lower numerosity, density, and contrast (the group of six oranges, left bars). However, extraction of mean size was not completely independent of these orthogonal visual cues; both children and adults correctly indicated which group had larger oranges more often when that group had higher numerosity, density, and contrast. * represents $p < .05$ and ** represents $p < .01$.

Experiment 1 could not have been due to numerosity, and is unlikely to have been due to density.

Discussion

We showed, for the first time, that ensemble-coding mechanisms are in place early in childhood. Even at 4–5 years of age, children were able to make precise and rapid comparisons of average size between groups. A conservative ideal-observer analysis revealed that children integrated information from more than four objects to estimate which group had larger objects overall. This value is nearly double the capacity of visual working memory at 3–5 years of age (Simmering, 2012), arguing against the possibility that children serially searched and compared the sizes of just a few random objects. A control experiment further ruled out this alternative hypothesis, showing that ensemble size integration proceeds even when children are unable to engage serial search. This integration process is independent of sensitivity to individual objects, and it even allowed children to extract average size information from groups despite poor sensitivity for perceiving individual objects. We also verified that this early summary-statistical pooling cannot be explained entirely by non-ensemble

information, like numerosity, density, or contrast. While surprisingly sensitive, children pooled information across fewer objects than adults, indicating that adults' remarkable ability to perceive the gist of groups and crowds develops gradually.

Our investigation provides a novel perspective on the sophistication of Gestalt perception in childhood. Previous investigations using hierarchical stimuli and coherent motion have shown that young children gradually develop sensitivity to global information in patterns and scenes (Dukette & Stiles, 2001; Elleberg, Lewis, Dirks, Maurer, Ledgeway, Guillemot & Lepore, 2004; Narasimhan & Giaschi, 2012; Parrish, Giaschi, Boden & Dougherty, 2005; Porporino, Iarocci, Shore & Burack, 2004; Prather & Bacon, 1986; Scherf, Berhmann, Kimchi & Luna, 2009). Even sensitivity to collective information, like numerosity, is in place in infancy (e.g. Zosh *et al.*, 2011). But perceiving groups – heterogeneous collectives of unique individuals – cannot occur by accessing global image attributes alone; it requires an extra step of averaging heterogeneous local information from many objects to extract a summary representation. Our findings demonstrate that ensemble coding allows children to do just this, even with some independence from numerosity, density, or contrast. Our results also lay the groundwork for future investigations to determine whether, like adults, young children's developing ensemble-coding mechanisms are flexible enough to enhance perception of complex social information in groups of people, like facial expression or crowd motion (Haberman & Whitney, 2007; Sweeny *et al.*, 2012, 2013). They also motivate future investigation on the relative role and prevalence of ensemble coding and individual object coding in young children's perception.

Perhaps most interestingly, our results are an important step in understanding the phenomenology of early visual experience. Many critical visual functions are still immature or limited at 4–5 years of age, including acuity and contrast sensitivity (e.g. Benedek *et al.*, 2003; Elleberg *et al.*, 1999; Jeon *et al.*, 2010; Skoczenski & Norcia, 2002), selective attention (e.g. Johnson, 2002), attentional resolution in space and time (Farzin *et al.*, 2010, 2011), visual working memory capacity (Cowan *et al.*, 2011; Simmering, 2012; Wilson *et al.*, 1987), and multiple object tracking (O'Hearn, Landau & Hoffman, 2005). Unchecked, these limitations would severely constrain the capacity of children's conscious experience, leaving cohesive groups to instead appear noisy and fragmented. However, we have shown that compressed summary representations may allow children to surmount these limitations and regard groups as ensembles. Just as prosopagnosics compensate for noisy perception

of individual faces by accessing the gist of an entire crowd (Yamanashi Leib *et al.*, 2012), our findings show that children have the requisite ensemble mechanisms that could, in principle, allow them to overcome their own perceptual limitations, grasping the broader gist of a scene instead of focusing on individual details.

Acknowledgements

We thank Alex McDonald for running experiments at the University of Denver. This study was supported by the National Institutes of Health grant R01 EY018216 and the National Science Foundation grant NSF 1245461.

References

- Alvarez, G.A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, **15** (3), 122–131. doi: 10.1016/j.tics.2011.01.003
- Alvarez, G.A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, **19** (4), 392–398. doi: 10.1111/j.1467-9280.2008.02098.x
- Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science*, **12** (2), 157–162.
- Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & Psychophysics*, **70**, 1325–1326.
- Awh, E., Barton, B., & Vogel, E.K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, **18** (7), 622–628.
- Benedek, G., Benedek, K., Kéri, S., & Janáky, M. (2003). The scotopic low-frequency spatial contrast sensitivity develops in children between the ages of 5 and 14 years. *Neuroscience Letters*, **345** (3), 161–164.
- Chong, S.C., Joo, S.J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: not so implausible after all. *Perception & Psychophysics*, **70**, 1327–1334.
- Chong, S.C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, **43** (4), 393–404.
- Cowan, N., AuBuchon, A.M., Gilchrist, A.L., Ricker, T., & Saults, J.S. (2011). Age differences in visual working memory capacity: not based on encoding limitations. *Developmental Science*, **14** (5), 1066–1074.
- Dakin, S.C., Bex, P.J., Cass, J.R., & Watt, R.J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision*, **9**, 1–16.
- Desimone, R., & Schein, S.J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, **57**, 835–868.
- Dukette, D., & Stiles, J. (2001). The effects of stimulus density on children's analysis of hierarchical patterns. *Developmental Science*, **4** (2), 233–251.

- Dumoulin, S.O., & Hess, R.F. (2007). Cortical specialization for concentric shape processing. *Vision Research*, **47**, 1608–1613.
- Elleberg, D., Lewis, T.L., Dirks, M., Maurer, D., Ledgeway, T., Guillemot, J.-P., & Lepore, F. (2004). Putting order into the development of sensitivity to global motion. *Vision Research*, **44** (20), 2403–2411.
- Elleberg, D., Lewis, T.L., Liu, C.H., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Research*, **39**, 2325–2333.
- Enns, J.T., & Girgus, J.S. (1985). Developmental changes in selective and integrative visual attention. *Journal of Experimental Child Psychology*, **40**, 319–337.
- Farzin, F., Rivera, S.M., & Whitney, D. (2010). Spatial resolution of conscious visual perception in infants. *Psychological Science*, **21**, 1502–1509.
- Farzin, F., Rivera, S.M., & Whitney, D. (2011). Time crawls: the temporal resolution of infants' visual attention. *Psychological Science*, **22** (8), 1004–1010.
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, **106**, 1389–1398.
- Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, USA*, **99**, 15822–15826.
- Goldberg, M.C., Maurer, D., & Lewis, L.L. (2001). Developmental changes in attention: the effects of endogenous cueing and of distractors. *Developmental Science*, **4**, 209–219.
- Gopnik, A. (2009). *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life*. New York: Picador/Farrar, Straus and Giroux.
- Gopnik, A., & Wellman, H.M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, **138** (6), 1085–1108. doi: 10.1037/a0028044 1085–1108
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, **17** (17), R751–R753. doi: 10.1016/j.cub.2007.06.039
- Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, **35** (3), 718–734. doi: 10.1037/a0013899
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, **18**, 855–859.
- Jeon, S.T., Hamid, J., Maurer, D., & Lewis, T.L. (2010). Developmental changes during childhood in single-letter acuity and its crowding by surrounding contours. *Journal of Experimental Child Psychology*, **107**, 423–437.
- Johnson, M.H. (2002). The development of visual attention: a cognitive neuroscience perspective. In T. Munakata & R.O. Gilmore (Eds.), *Brain development and cognition: A reader* (pp. 134–150). Oxford: Blackwell.
- Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition*, **83**, B35–B42.
- Koffka, K. (1935). *The principles of gestalt psychology*. London: Routledge and Kegan Paul.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, **390** (6657), 279–281. doi: 10.1038/36846
- Myczek, K., & Simons, D.J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, **70** (5), 772–788.
- Narasimhan, S., & Giaschi, D.E. (2012). The effect of dot speed and density on the development of global motion perception. *Vision Research*, **62** (1), 102–107.
- O'Hearn, K., Landau, B., & Hoffman, J.E. (2005). Multiple object tracking in people with William's syndrome and in normally developing children. *Psychological Science*, **16** (11), 905–912.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, **4** (7), 739–744. doi: 10.1038/89532
- Parrish, E.E., Giaschi, D.E., Boden, C., & Dougherty, R. (2005). The maturation of form and motion perception in school age children. *Vision Research*, **45** (7), 827–837.
- Porporino, M., Iarocci, G., Shore, D.I., & Burack, J.A. (2004). A developmental change in selective attention and global form perception. *International Journal of Behavioral Development*, **28** (4), 358–364.
- Prather, P.A., & Bacon, J. (1986). Developmental differences in part/whole identification. *Child Development*, **57** (3), 549–588.
- Proulx, M.J. (2010). Size matters: large objects capture attention in visual search. *PLoS ONE*, **5**, e15293. doi: 10.1371/journal.pone.0015293
- Ristic, J., & Kingstone, A. (2009). Rethinking attentional development: reflexive and volitional orienting in children and adults. *Developmental Science*, **12**, 289–296.
- Scherf, K.S., Berhmann, M., Kimchi, R., & Luna, B. (2009). Emergence of global shape processing continues through adolescence. *Child Development*, **80** (1), 162–177.
- Scholl, B.J., & Pylyshyn, Z.W. (1999). Tracking multiple items through occlusion: clues to visual objecthood. *Cognitive Psychology*, **38** (2), 259–290.
- Semenov, L.A., Chernova, N.D., & Bondarko, V.M. (2000). Measurement of visual acuity and crowding effect in 3–9-year-old children. *Human Physiology*, **26** (1), 16–20.
- Simmering, V.R. (2012). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology*, **111**, 695–707.
- Simons, D.J., & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Sciences*, **1** (7), 261–267. doi: 10.1016/S1364-6613(97)01080-2
- Skoczenski, A.M., & Norcia, A.M. (2002). Late maturation of visual hyperacuity. *Psychological Science*, **13** (6), 537–541.
- Spelke, E.S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, **99** (4), 605–632.
- Sweeny, T.D., Haroz, S., & Whitney, D. (2012). Reference repulsion in the categorical perception of biological motion. *Vision Research*, **64**, 26–34.
- Sweeny, T.D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: sensitive ensemble coding mechanisms for

- biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, **39** (2), 329–337.
- Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology*, **78**, 392–402.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, **95**, 15–48.
- Whitney, D., Haberman, J., & Sweeny, T.D. (2014). From textures to crowds: multiple levels of summary statistical perception. In J.S. Werner & L.M. Chalupa (Eds.), *The new visual neurosciences*. Cambridge, MA: MIT Press.
- Whitney, D., & Levi, D.M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, **15** (4), 160–168.
- Wilson, J.T., Scott, J.H., & Power, K.G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, **5**, 249–255.
- Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, USA*, **105**, 5012–5015.
- Yamanashi Leib, A., Puri, A.M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia*, **60** (7), 1698–1707.
- Zosh, J.M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, **140**, 141–158.

Received: 16 April 2013

Accepted: 14 July 2014