Developmental Science

Developmental Science 14:1 (2011), pp 125–135 DOI: 10.1111/j.1467-7687.2010.00962.x

PAPER

The development of numerical estimation: evidence against a representational shift

Hilary C. Barth and Annie M. Paladino

Department of Psychology, Wesleyan University, USA

Abstract

How do our mental representations of number change over development? The dominant view holds that children (and adults) possess multiple representations of number, and that age and experience lead to a shift from greater reliance upon logarithmically organized number representations to greater reliance upon more accurate, linear representations. Here we present a new theoretically motivated and empirically supported account of the development of numerical estimation, based on the idea that number-line estimation tasks entail judgments of proportion. We extend existing models of perceptual proportion judgment to the case of abstract numerical magnitude. Two experiments provide support for these models; three likely sources of developmental change in children's estimation performance are identified and discussed. This work demonstrates that proportion-judgment models provide a unified account of estimation patterns that have previously been explained in terms of a developmental shift from logarithmic to linear representations of number.

Introduction

How do mental representations of number change throughout development? The most prominent account holds that children possess multiple mental representations of number: they initially rely on less accurate logarithmic number representations, becoming more likely to rely on more accurate linear representations with age and experience (Booth & Siegler, 2006, 2008; Laski & Siegler, 2007; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003). Findings said to support the logarithmic-to-linear-shift account come largely from number-line estimation tasks, in which participants estimate the magnitudes of presented numerical quantities by marking their proper positions on number lines. When participants' estimates are plotted with respect to the true presented quantities, accurate estimates lie on the line $y = x$. For a given number line, younger children tend to produce less accurate estimates that are fit better by a logarithmic curve than by a line, and older children tend to produce more accurate, more nearly linear estimates. This general pattern of results has been replicated across many studies and it appears quite robust, but it is not clear that the idea of a logarithmic-to-linear representational shift provides the best explanation of these results. In the present paper, we present an alternative account based on proportion judgments.

An array of important theoretical and practical issues turns on the interpretation of results from number-line estimation tasks. The logarithmic-to-linear representational shift account has inspired investigations of educational interventions meant to improve math performance by encouraging children to use linear rather than logarithmic representations (Siegler & Ramani, 2008, in press). Performance patterns thought to indicate linear representation use are correlated with performance on standardized math achievement tests in kindergarteners through fourth graders, and it has been argued that this representational shift may play a critical role (Siegler & Ramani, 2008; Siegler, Thompson & Opfer, in press; Siegler & Ramani, in press). In addition, the log-to-linear-shift framework has been used to characterize elements of mathematical learning disability (Geary, Hoard, Byrd-Craven, Nugent & Numtee, 2007; Geary, Hoard, Nugent & Byrd-Craven, 2008). It has even been invoked in cross-cultural studies exploring the ultimate origins of mathematics (Dehaene, Izard, Spelke & Pica, 2008).

There are compelling reasons, however, to question the idea that children's number-line estimation performance provides evidence of a representational shift. First, if a general shift in children's mental representations of number caused the observed change from more logarithmic-looking to more linear-looking estimates, then other estimation tasks should produce comparable results in similar populations. Yet this pattern of findings does not necessarily appear in estimation tasks without number lines (e.g. Barth, Starr & Sullivan, 2009; Lipton

Address for correspondence: Hilary Barth, Department of Psychology, Wesleyan University, 207 High St, Middletown, CT 06459, USA; e-mail: hbarth@wesleyan.edu

& Spelke, 2005; see also Noël, Rouselle & Mussolin, 2005). Second, the 'linear' responses of older children in number-line tasks may not be so linear: these children exhibit systematic patterns of over- and underestimation (e.g. Siegler & Booth, 2004; Booth & Siegler, 2006). These patterns remain unexplained by the log-to-linear-shift hypothesis, but they are predicted by the alternative explanation proposed here. Third, the logarithmic-to-linear-shift account does not recognize a critical property of number-line estimation tasks: that they are proportion judgments. For example, in a number-line task with Arabic numerals, children might be asked to mark the position of '30' on a line with endpoints labeled '0' and '100'. To perform this task, they cannot simply estimate the numerical magnitude of '30' in isolation; rather, they must estimate the size of a part (the numerical magnitude of '30') relative to the size of the whole (the magnitude of '100'). This point about proportion judgments forms the foundation of our alternative account. The fact that number-line estimation requires proportion judgments leads to specific, yet untested, predictions about performance patterns that are consistent with the systematic patterns of over- and underestimation that appear in children's 'linear' estimates. The present paper tests these predictions, and argues that a proportion-judgment account of the development of numerical estimation provides a better explanation than does the idea of a logarithmic-to-linear representational shift.

Proportion judgments have been thoroughly studied and modeled in many domains, sometimes with paradigms that are quite similar to number-line estimation tasks. In one study, for example, fourth grade students completed an explicit proportion-judgment task in which they observed a pie chart divided into two sections, and were asked to divide a horizontal line into two parts to match the proportion of the sections (Spence & Krizel, 1994). In perceptual proportion-judgment tasks like this one, participants observe the part and the whole directly to assess their magnitudes. A number-line task with Arabic numerals is more abstract: participants in such a task must recall the proper magnitudes associated with the presented numerals. Despite the greater degree of abstraction required by number-line tasks, however, quantitative predictions developed for perceptual proportion judgments apply to number-line estimation as well. Here, we extend existing explanations of perceptual proportion judgment (Spence, 1990; Hollands & Dyre, 2000) to the abstract continuum of numerical magnitude.

The outcome of a proportion judgment depends on the biases involved in estimating the individual part and whole magnitudes (Spence, 1990). Often, 'psychological' (or 'subjective') magnitude appears to be a compressed function of actual stimulus magnitude. Relationships between psychological and actual magnitudes are often well described by negatively accelerating power functions, with exponents between 0 and 1 (for some sensory continua, these exponents are greater than 1, but they will not be dealt with here). For example, the estimated area of a circle does not increase as a linear function of the circle's physical area; instead, their relationship is generally well described by a power function with an exponent of 0.8 (e.g. Teghtsoonian, 1965; Chong & Treisman, 2003), such that estimates of area are systematically biased. This relationship between psychological and actual magnitude holds not only for many perceptual continua, but also for more abstract numerical magnitude (e.g. Nieder & Miller, 2003; Izard & Dehaene, 2008). When a judgment of *proportion* is made, two magnitudes must be estimated rather than just one, so estimates of proportion are no longer well described by single power functions.

Because of the consequences of making biased estimates of two magnitudes (both the part and the whole) in a proportion judgment, estimates of proportion often take the form shown schematically in Figure 1A (see

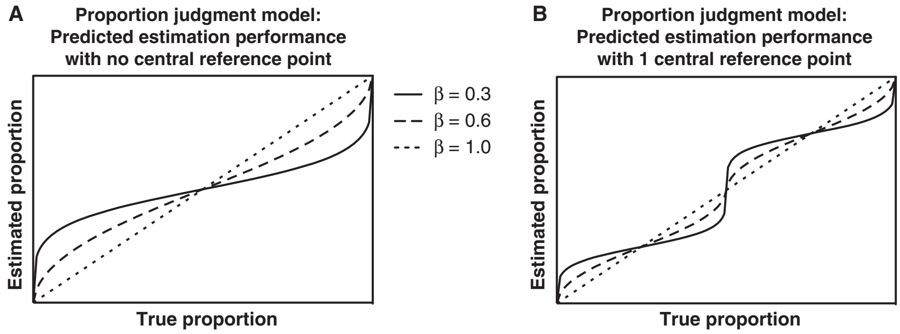

Figure 1 *A. Predicted proportion estimation data when observers judge the magnitude of the part relative to the entire whole (basic 1-cycle model; Spence, 1990): proportions less than 0.5 are overestimated, and proportions greater than 0.5 are underestimated. B. Predicted form of proportion estimation data when observers use one central reference point (2-cycle model, with overestimation of proportions less than 0.25, underestimation from 0.25 to 0.5, overestimation from 0.5 to 0.75, underestimation from 0.75 to 1.0; and a steep crossing of the y = x line at (50, 50); Hollands & Dyre, 2000). In both cases, β is the exponent in a power function describing the relationship of psychological to physical magnitude. For equal values of β, observers who use a central reference point produce more accurate (more nearly linear) estimates than those who don't. When β = 1, y = x.*

Spence, 1990; Hollands & Dyre, 2000, for proofs). This *power model of proportion judgment* (Spence, 1990) provides an excellent explanation of adults' performance on a variety of perceptual proportion tasks (e.g. Shuford, 1961; Nakajima, 1987; Varey, Mellers & Birnbaum, 1990; as demonstrated by Hollands & Dyre, 2000). It is also surprisingly simple, requiring only one free parameter, β: the exponent determining the shape of the power function relating psychological to actual magnitude. This model predicts overestimation of small proportions (below 0.5) and underestimation of large proportions (above 0.5), with little to no bias at 0.5 (that is, the estimation function should pass through the line $y = x$ at around that point), as depicted in Figure 1A. Figure 1A also illustrates how the proportion judgments required by number-line estimation tasks could produce deceptively logarithmic-looking estimation data in younger children, particularly when the upper end of the range is sparsely sampled as it was in previous studies (e.g. Booth & Siegler, 2006). The farther β is from 1, the greater the potential resemblance to a logarithmic estimation pattern – and for perceptual proportion judgments, β values are farther from 1 for younger children (Spence & Krizel, 1994; Hollands & Dyre, 2000).

The simple power model of proportion judgment described above (Figure 1A; Spence, 1990) applies when participants judge parts relative to entire wholes. But observers may make use of intermediate reference points as well, a strategy that should lead to different performance patterns. Spence's (1990) model was generalized to account for such cases by later researchers, who developed a *cyclical power model* (Hollands & Dyre, 2000). The cyclical model explained adults' performance in many perceptual proportion tasks (e.g. Huttenlocher, Hedges & Duncan, 1991) as well as children's estimates of proportion in graphs (Spence & Krizel, 1994). Because the design of number-line estimation tasks frequently provides participants with convenient central reference points (for example, by giving children an example trial in which they are shown where '50' goes in a 0–100 number-line task), we will be particularly concerned with performance patterns predicted when observers judge part magnitudes relative to central reference points (Figure 1B; see Experiment 1). This model predicts overestimation below 0.25, underestimation between 0.25 and 0.5, overestimation between 0.5 and 0.75, and underestimation above 0.75. It also predicts smooth crossings of the line $y = x$ at 0.25 and 0.75, with a steep crossing at 0.5 (Hollands & Dyre, 2000; Figure 1B).

In two experiments, we test the predictions of models of perceptual proportion judgment (Spence, 1990; Hollands & Dyre, 2000) in the context of children's number-line estimation. This work argues against the idea of a developmental shift from logarithmic to linear mental representations of number, demonstrating instead that the proportion-judgment account provides a unified, theoretically motivated explanation of numerical estimation performance, both in older children who produce generally linear-appearing data and in younger children who produce generally logarithmic-appearing data for the same number-line task.

Experiment 1

In Experiment 1, we tested the predictions of the proportion-judgment account for numerical estimation data that have previously been explained in terms of linear mental representations of number. We presented children averaging 7 years of age with a number-line task that was age appropriate for producing linear estimation patterns (using a 0–100 number line). Studies of number-line estimation typically oversample at the low end of the range in order to distinguish fits of logarithmic vs. linear models of performance (but see Ebersbach, Luwel, Frick, Onghena & Verschaffel, 2008). Because proportion-judgment models make specific predictions about responses at the upper end of the range as well, we adjusted our sampling method accordingly.

We tested two power models of proportion judgment, each with only one free parameter (β, the exponent determining the shape of the power function relating psychological to physical magnitude). We tested the one-cycle version of the model (Figure 1A; Spence, 1990), which assumes that children judge the size of the given numeral relative to the size of the whole ('100'). We also tested a cyclical model with two cycles (Figure 1B; Hollands & Dyre, 2000), which applies to proportion judgments in which a central reference point is used. Because the numerals to be estimated were printed over the center of the number line and children were first shown the proper location of '50' (as in Booth & Siegler, 2006), it is likely that some children will use a central reference point; this should yield more accurate estimates on average (see Figure 1B). Both models predict alternating patterns of over- and underestimation (see Figures 1A and 1B), and both predict that estimation functions should pass through the points (0,0), (50, 50), and (100, 100), deviating from $y = x$ between these points (Spence, 1990; Hollands & Dyre, 2000). The two-cycle model predicts that estimation functions should also pass through (25, 25) and (75, 75) (Hollands & Dyre, 2000). Both models reduce to $y = x$ if the parameter β is equal to 1; therefore, either model can successfully fit accurate (near-linear) estimates.

Method

Twenty-one children (mean age 89 months, range 78–102 months) were recruited from local families. The number-line task was adapted from Booth and Siegler (2006). Stimuli consisted of a booklet with a 23-cm number line on each page, marked with '0' at the left end and '100' at the right end, with the numeral to be estimated printed 2.5 cm above the center. For an initial practice trial, there were several number lines marked only with ‘0’ and ‘100’. Two sets of trials were presented to each child in order to sample thoroughly at both low and high ends of the range. The ‘small’ set contained the 26 numerals presented by Booth and Siegler: 3, 4, 6, 8, 12, 14, 17, 18, 21, 24, 25, 29, 33, 39, 42, 48, 52, 57, 61, 64, 72, 79, 81, 84, 90, 96. The ‘large’ set contained 26 numbers generated by subtracting each ‘small’ numerosity from 100: 4, 10, 16, 19, 21, 28, 36, 39, 43, 48, 52, 58, 61, 67, 71, 75, 76, 79, 82, 83, 86, 88, 92, 94, 96, 97. Approximately half the children received the ‘small’ set first and half received the ‘large’ set first (52 trials per child). Within each set, trial order was randomized for each participant.

The experimenter said, ‘We’re going to play a game with number lines. I’ll ask you to show me where you think some numbers should go on the number line. When you decide where the number goes, make a line through the number line like this.’ The experimenter then marked a sample number line. The experimenter then prompted participants to try marking a new number line to show where ‘50’ should go. After responding, participants were shown a number line marked in the middle. The experimenter asked if they knew why 50 went there and explained, ‘because 50 is half of 100, it goes right in the middle between 0 and 100. So 50 goes right there, but it’s the only number that goes there.’ On the first test trial, if participants marked the halfway point, the experimenter said that only 50 goes in the middle. For each trial, the experimenter asked, ‘where would you put ___?’ and said, ‘good job’ or ‘thank you’ after participants responded. Marked positions were translated into numerical estimates (Booth & Siegler, 2006).

Results

Figure 2 depicts median estimates for this Age 7 group; Figure 3 depicts individual children’s estimates. Both figures show nonlinear fits for the better-fitting of the two simple proportion-judgment models described earlier: the

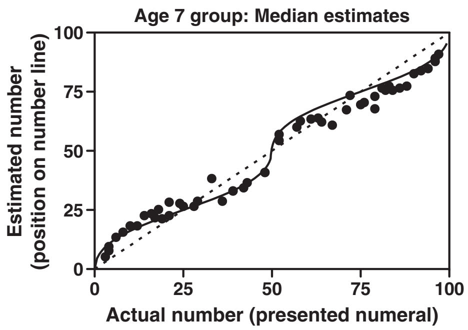

Figure 2 *Median estimates for the Age 7 group, Experiment 1. The solid line shows the preferred (2-cycle) model, and the dotted line shows y = x. Two participants who placed at least half of their estimates at the midpoint of the number line and/or produced estimates unrelated to the presented numeral were not included in the depicted medians (analyses were conducted both with and without these children:* R^2 *= 0.9798 without the two excluded children,* R^2 *= 0.9735 with all included).*

one-cycle model (Spence, 1990; see Table S1) and the two-cycle model resulting when observers use central reference points (Hollands & Dyre, 2000; Table S1). Likelihood ratios, which indicate the explanatory power of a particular model, were used to assess the better model of performance for each participant; this method allows us to compare models with different numbers of parameters, which cannot be done by comparing R^2 values alone (Glover & Dixon, 2004; Burnham & Anderson, 2002). Table 1 gives the results of these comparisons (including R^2 values) for each child (excluding two children who marked the center of the number line on at least half of the trials and/or produced estimates unrelated to the presented numeral; see Figure 3). The two-cycle proportion judgment model provided the better explanation of this group’s median estimates (Figure 2), and of the individual estimates of 8/21 children (Table 1; Figure 3).

We also considered the fit of the linear model favored by the log-to-linear-shift hypothesis. This model has two parameters (a slope and a y-intercept). The proportion-judgment model provided a better explanation of every child’s individual estimates and a better explanation of the group median data. We then considered a simpler linear model with only a single parameter, the slope (with a y-intercept of zero). Again the proportion-judgment model (also a single-parameter model) provided a better explanation of 7-year-olds’ group medians and individual data, with the one-parameter linear model providing poorer fits for every individual child except two (S36 and S47; Figure 3).

Discussion

Proportion-judgment models provided an excellent explanation of number-line estimation performance in the Age 7 group, both for individual children’s estimates and for the group medians. The proportion-judgment model provided a better explanation of performance than did the two-parameter linear model favored in the logarithmic-to-linear-shift account. The proportion-judgment model also provided a better explanation than did a simpler one-parameter linear model. Understanding this task as a proportion judgment provides an explanation of cyclical patterns of bias in children’s estimates, as other accounts cannot.

Children in this age range typically produce more linear-looking than logarithmic-looking responses for a 0–100 number line. Can the proportion-judgment account also explain the typical logarithmic-looking performance patterns of younger children on the same number-line task? Experiment 2 explores the proportion-judgment model in the context of younger children’s estimates.

Experiment 2

In Experiment 2, we tested the predictions of the proportion-judgment account for number-line estimation

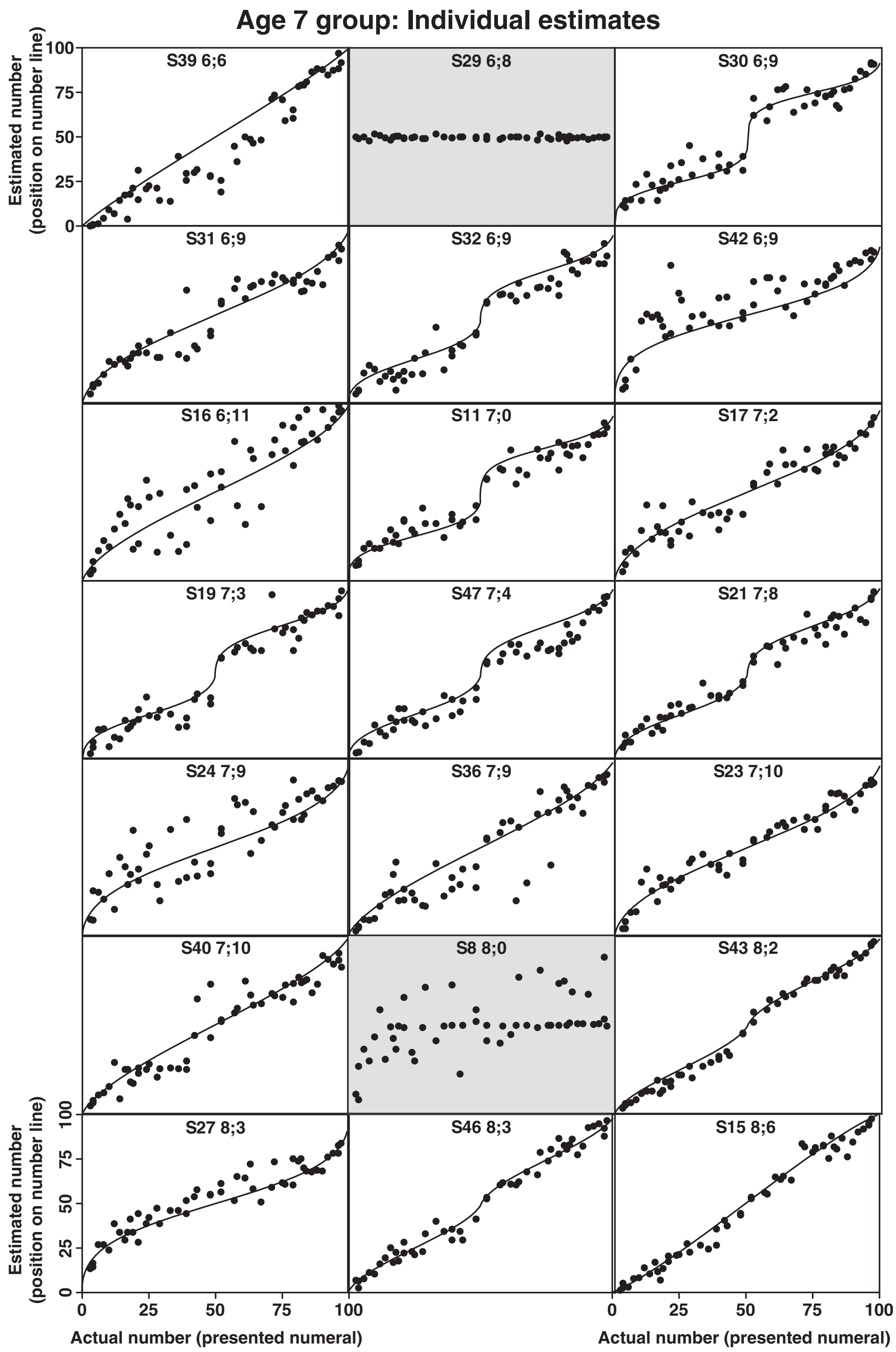

Figure 3 *Estimates of individual participants from the Age 7 group, Experiment 1, arranged in order of age. Solid lines show the preferred of two tested models: 1-cycle proportion model or 2-cycle proportion model. Shaded plots without fits indicate participants who placed at least half of their estimates at the midpoint of the number line and/or produced estimates unrelated to the presented numeral.*

data that have previously been explained in terms of children's logarithmic mental representations of number. Experiment 1 showed that existing single-parameter versions of the proportion-judgment model (Figure 1A, 1B; Spence, 1990; Hollands & Dyre, 2000) provide an excellent explanation of performance for the Age 7 group. However, as one can see by comparing the predictions of existing proportion-judgment models with previous estimation data from younger children, these models must necessarily encounter difficulty when confronted with younger children's estimates. This is because these single-parameter models assume that observers judge each given part relative to a known whole (Spence, 1990; Hollands & Dyre, 2000). This assumption has an important practical consequence with respect to possible shapes that the model can take: fits of the single-parameter models tested in Experiment 1 must pass through (0, 0), (50, 50), and (100, 100).

The assumption described above makes sense for perceptual proportion judgments, in which the part and the

Table 1 *Experiment 1, fits of proportion judgment models to data from individual children in the Age 7 group. For each child, the table lists the proportion judgment model that was better supported by the data (1-cycle or 2-cycle), and values of* R^2 *and parameter* β *(the exponent determining the shape of the power function relating psychological to physical magnitude) for the better-supported model*

Age (mo)	S #	Prop-1 or Prop-2?	R^2	β
78	39	1	.9241	0.8572
80	29			
81	30	2	.9440	0.3713
81	31	1	.8995	0.6412
81	32	2	.9376	0.4914
81	42	1	.6126	0.4050
83	16	1	.7844	0.7273
84	11	2	.9371	0.4071
86	17	1	.9080	0.6188
87	19	2	.9472	0.4875
88	47	2	.9163	0.5307
92	21	2	.9619	0.5550
93	24	1	.7589	0.5421
93	36	1	.8249	0.7544
94	23	1	.9471	0.6245
94	40	1	.9105	0.7780
96	8			
98	43	2	.9875	0.7488
99	27	1	.8893	0.4540
99	46	2	.9799	0.7462
102	15	1	.9778	1.125

whole are perceptually available to observers (Spence, 1990; Hollands & Dyre, 2000). But it should not necessarily apply for young children's proportion judgments with Arabic numerals: in our task, the 'whole' is the magnitude of the numeral '100'. The assumption should be valid for most children in our Age 7 group (consistent with their estimation patterns), who are in an age group that is typically knowledgeable about numbers up to 100. It will certainly not be valid for all of the younger children who produce logarithmic-looking estimates in similar tasks: 5-year-olds, for example, exhibit a wide range of levels of knowledge of numbers from 1 to 100 (e.g. Barth *et al.*, 2009; Ebersbach *et al.*, 2008; Lipton & Spelke, 2005). Therefore, we should expect that for at least some children in this age range, the assumptions of the original single-parameter proportion-judgment models should be violated, and estimates should not be fit well by a curve passing through the points (0, 0), (50, 50), and (100, 100). Previous findings are consistent with this prediction (e.g. Siegler & Opfer, 2003).

Instead of appealing to logarithmic number representations to account for young children's estimates, we propose that the idea of a proportion judgment relative to an unknown or uncertain whole magnitude provides a better explanation. Here we introduce adapted versions of the proportion models tested in Experiment 1 (see Table S1). The original models assume that the observer knows something about the magnitude of the whole (i.e. has associated an appropriate mental magnitude with '100'), and judges the presented numeral relative to that known whole magnitude. The adapted versions include a second parameter, *W*, to account for young children's lack of knowledge of the magnitude of '100'. The parameter *W* can be thought of as the *de facto* numerical magnitude of the whole in the part/whole judgment:[1] in these adapted models, the magnitude of the whole is a parameter to be fitted, rather than a known value fixed at 100.[2] Of course these models do not capture all potential sources of variability in these children's estimates: for example, children ignorant of the magnitude of '100' are unlikely to judge the presented numerals relative to a single, stable whole magnitude across all trials, and they are likely to be unfamiliar with the magnitudes of other presented numerals as well. Fits of the modified models must pass through (0,0), but are no longer forced through (50,50) and (100,100). These adapted models have only two free parameters, like logarithmic or linear models of estimation.

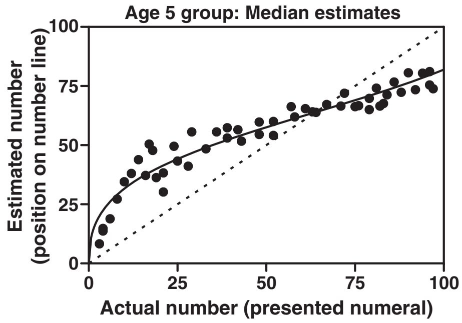

Figure 4 *Median estimates for the Age 5 group, Experiment 2. The solid line shows the preferred model (the adapted 1-cycle proportion model). The dotted line shows y = x. Five participants who placed at least half of their estimates at the midpoint of the number line and/or produced estimates unrelated to the presented numeral were not included in the depicted medians (analyses were conducted both with and without these children:* R^2 *= 0.9216 without the excluded children,* R^2 *= 0.8799 with all included).*

We presented children averaging 5 years of age with a number-line task that was age appropriate for producing logarithmic-appearing estimates. As in Experiment 1, children were presented with a number line up to 100.

Method

Twenty-one children (mean age 67 months, range 61–73 months) participated. Testing was carried out as in Experiment 1: children in the Age 5 group completed the same 0–100 number-line task.

[1] For example, a *W* value of 120 means that a child appears to be using a numerical magnitude of about 120, rather than 100, as the whole magnitude in his/her proportion judgments, on average. This suggests that the child has *underestimated* the magnitude of the numeral '100' (a *W* value of 120 would mean that the child had mapped a magnitude of about 83, on average, onto the numeral '100'). The parameter *W* does *not* equal the magnitude associated with the numeral '100'.

[2] This idea is distinct from the hypothesis, tested by Siegler and Opfer (2003), that children simply treat the task as an open-ended estimation.

Results

Median estimates for the Age 5 group are shown in Figure 4, and individual children's estimates are in Figure 5. Median and individual estimates were fitted by the two adapted two-parameter proportion-judgment models: a one-cycle model and a two-cycle model (Table S1). In both figures, solid lines indicate nonlinear fits for the better-fitting of the two models. Models were compared as in Experiment 1; Table 2 gives results for each child. The one-cycle model provided the better explanation of this group's median estimates (see Figure 4) and for the individual estimates of 19/21 children (Table 2 and Figure 5).

R^2 values for the adapted proportion-judgment model fits were compared to R^2 values for logarithmic fits to individual children's data. Each of these models requires two parameters, so the R^2 values associated with individual children's estimates for each model may be compared directly (Opfer, nd). The means of the individual children's R^2 values were 0.5369 and 0.5376, respectively; both sets of R^2 values passed a D'Agostino and Pearson normality test. These were not significantly different: $t(15) = 0.04$, $p > .05$.

Discussion

The adapted proportion-judgment models provided a good explanation of 5-year-olds' estimates in the 0–100 number-line task. For some 5-year-olds, the fitted value of *W* was close to 100, suggesting that they had used an appropriate whole magnitude in their numerical proportion judgments. For others, who produced data that clearly could not be fit well by simpler models that must pass through (50,50) and (100,100), the adapted model also provided a good explanation of the data, with fitted values of *W* somewhat larger than 100 (corresponding to a tendency to underestimate the magnitude of the numeral '100'). These results are consistent with our predictions, as 5-year-olds exhibit a wide range of skill levels at mapping numbers near 100 to their appropriate numerical magnitudes (e.g. Lipton & Spelke, 2005; Barth *et al*., 2009).

We also tested the adapted models on the Age 7 group's data from Experiment 1, predicting that their greater knowledge of the relevant numerical range would produce *W* values near 100. The mean value of *W* for older children was 108.9 (SD = 15.5), consistent with older children's greater knowledge of the mappings between numbers from 0 to 100 and their associated magnitudes.

Both the adapted proportion-judgment models introduced here and the logarithmic model favored by the logarithmic-to-linear-shift account provided equally good explanations of 5-year-old children's performance. Proportion-judgment models fit 5-year-olds' estimates equally as well as logarithmic models. The logarithmic-to-linear-shift account, however, requires two distinct two-parameter models to provide a reasonable explanation of both younger children's and older children's data. Our two-parameter proportion-judgment model explains both younger and older children's estimation patterns.

Table 2 *Experiment 2, fits of proportion judgment models to data from individual children in the Age 5 group. For each child, the table gives the proportion judgment model that was better supported by the data (adapted 1-cycle model or adapted 2-cycle model), and values of* R^2 *and parameters β (the exponent determining the shape of the power function relating psychological to physical magnitude) and* W *(the fitted magnitude of the whole in the part/whole judgment) for the better-supported model*

Age (mo)	S #	Prop-1A or Prop-2A?	R^2	β	*W*
61	7				
61	37				
61	48				
63	50	1	0.5865	0.3806	133.60
63	51	1	0.2282	0.2048	100.10
65	12	2	0.8207	0.5442	99.90
65	26	1	0.6444	0.4297	136.90
66	13	1	0.7863	0.5851	97.08
66	28	1	0.4510	0.2602	192.80
67	22				
69	3	1	0.4003	0.3315	127.50
69	25	1	0.5493	0.2435	97.03
69	38	1	0.2659	0.2695	112.30
70	34	1	0.1612	0.2131	120.50
70	44	1	0.2077	0.2168	146.20
70	33				
71	9	1	0.3777	0.3576	129.10
71	35	1	0.8881	0.6733	115.50
71	45	1	0.4828	0.3104	102.90
71	49	1	0.9034	0.6453	129.40
73	41	2	0.8365	0.4667	104.00

General discussion

We propose that because proportion judgments are required in number-line estimation tasks, number-line estimation should be interpreted within the theoretical framework of proportion judgments. A unified theoretical account based on proportion judgment provides a better explanation of the development of number-line estimation than does the logarithmic-to-linear representational shift account. The log-to-linear-shift account makes use of two distinct two-parameter models, yet it does not provide a comparably strong explanation of older children's estimates. The proportion-judgment account, making use of a single two-parameter model, provides an equally good explanation of younger children's estimates when compared to a logarithmic model, and a better explanation of older children's estimates when compared to a linear model.

Two experiments tested the predictions of our account of children's estimation in number-line tasks. In Experiment 1, simple existing models of proportion judgment, developed for perceptual tasks (Spence, 1990; Hollands & Dyre, 2000), provided an excellent explanation of 7-year-olds' performance on a 0–100 number-line

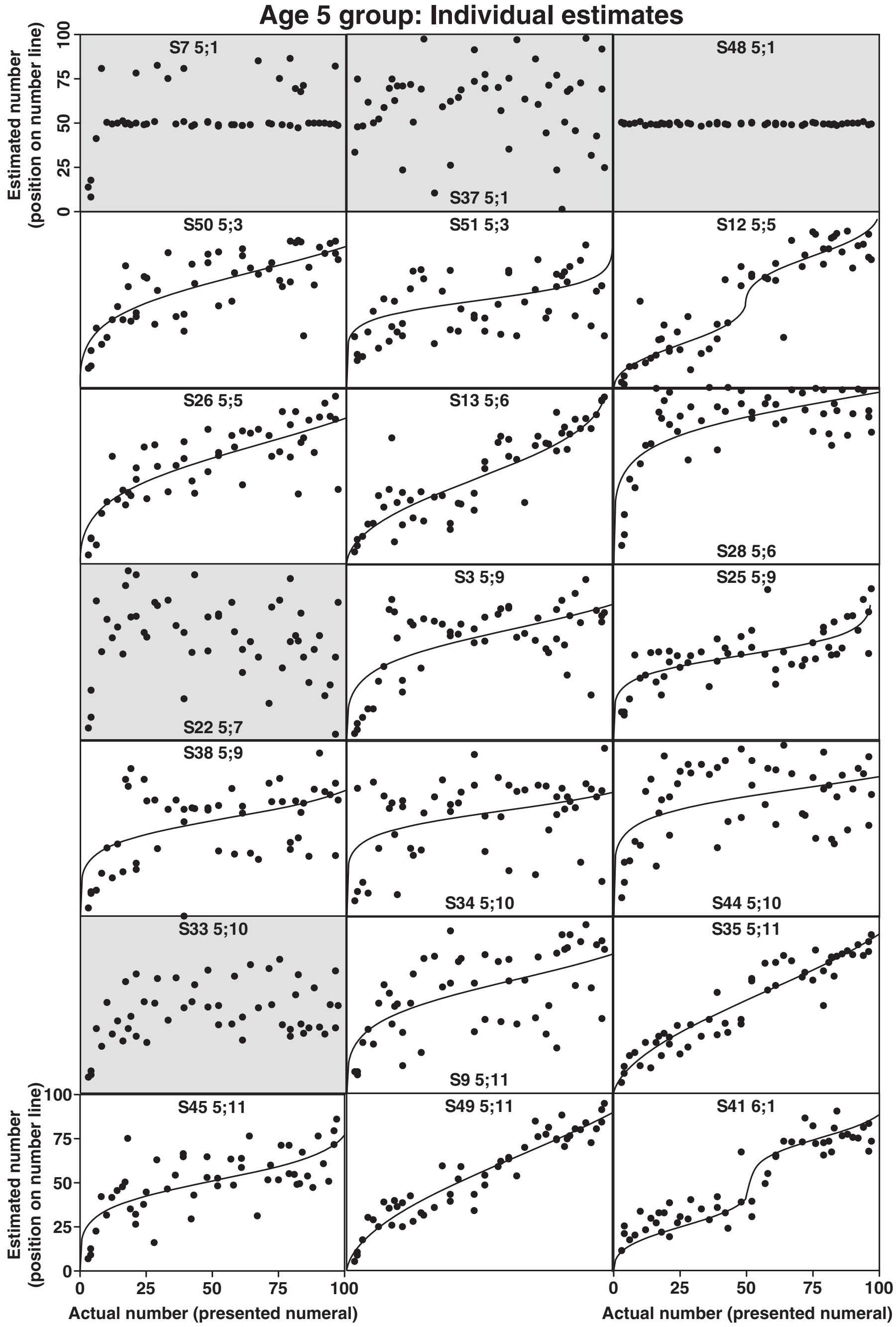

Figure 5 *Estimates of individual participants from the Age 5 group, Experiment 2, arranged in order of age. Solid lines show the preferred of the tested models (adapted 1-cycle proportion model or adapted 2-cycle proportion model). Shaded plots without fits indicate participants who placed at least half of their estimates at the midpoint of the number line, and/or produced estimates unrelated to the presented numeral.*

estimation task, performing better than linear models. These models account for patterns of cyclical bias in children's estimates that are not explained by the logarithmic-to-linear-shift hypothesis, or by other accounts of children's numerical estimation (e.g. Ebersbach *et al.*, 2008; Moeller, Pixner, Kaufmann & Nuerk, 2009).

Existing proportion-judgment models assume that observers have some knowledge of the magnitudes used in the task, because they were developed for tasks in which those magnitudes could be perceptually accessed. This assumption should be violated when 5-year-olds attempt a 0–100 number-line task, because many children of this age have not yet acquired reasonable mappings between numerals as large as 100 and their associated nonverbal numerical magnitudes (e.g. Lipton & Spelke, 2005; Barth *et al.*, 2009). For this reason, in Experiment 2 we introduced an adapted proportion-judgment model that allows for uncertainty in the magnitude of the whole in the part/whole judgment. This adapted model, requiring only two free parameters, provided a good explanation of 5-year-olds' performance on the 0–100 number-line task (as well as 7-year-olds').

One argument that has been presented in support of the log-to-linear-shift hypothesis comes from the finding that the same children may produce linear-looking estimates for one numerical range, and logarithmic-looking estimates for a larger range (e.g. Siegler & Opfer, 2003). This finding has been thought to suggest that children possess multiple numerical representations, with choices among them depending on age and experience. In our view, it is incorrect to suppose that children invoke logarithmic mental number representations in some contexts and linear representations in others. Instead, numerical proportion judgments for well-known ranges are made relative to known whole magnitudes (as in the case of the 7-year-olds in Experiment 1), but proportion judgments for less well-known ranges may be made relative to uncertain wholes (as in the case of most 5-year-olds in Experiment 2). The developmental change in children's estimation patterns, then, provides no evidence of a shift in the type of mental number representation used. The very same representation – a power-function representation of number – can result in drastically different patterns of performance when it is invoked in a proportion judgment over a well-known vs. a less-known numerical range.

Other models of number-line estimation

The logarithmic-to-linear-shift hypothesis has been questioned by other researchers studying number-line estimation as well (Ebersbach *et al.*, 2008; Moeller *et al.*, 2009). According to both of these accounts, the logarithmic-looking estimation patterns of relatively younger children are due to distinct linear representations of smaller vs. larger numbers. Ebersbach and colleagues (2008) proposed that children's estimates might be best understood in terms of a segmented linear model. According to this view, the mental number lines of relatively younger children are not logarithmically organized; rather, they are linear with a steep slope for familiar numbers, and linear with a shallower slope for unfamiliar numbers. Relatively older children therefore might produce linear-looking estimates on the same number-line task because they are generally familiar with all of the relevant numbers. This familiarity-based account makes use of a model with four parameters (the y-intercept of the first linear segment, the slopes of both segments, and the transition point between the two segments that should correspond to the transition between familiar and unfamiliar numbers). Moeller and colleagues also proposed a segmented linear model of children's number-line estimation, suggesting that young children may possess separate representations of one- and two-digit numbers that have not yet been integrated successfully. On this view, young children have a linear representation of one-digit and two-digit numbers, but the representation of two-digit numbers has a shallower slope. This decade-change account makes use of a model with three parameters (the y-intercept of the first linear segment and the slopes of both segments, with the transition between the two at a fixed point).

Both of these accounts based on segmented linear models require more parameters than the logarithmic and linear models employed by the log-to-linear-shift hypothesis. They also require more parameters than the proportion-judgment account proposed here. For this reason, they are unlikely to prevail in the model selection process unless they yield fits that are substantially better than those yielded by simpler models (e.g. Burnham & Anderson, 2002; Glover & Dixon, 2004). More complex models may also be favored if they are backed by strong theoretical motivation; however, no such motivation exists for these segmented linear models (Dehaene *et al.*, 2008). We did, however, consider these accounts by fitting both three- and four-parameter versions of a segmented linear model to our 5-year-olds' estimates, finding that they failed to provide competitive explanations of the data for group medians and for individual estimates.

The proportion-judgment account: potential sources of change

The proportion-judgment account proposed here identifies multiple sources of developmental change in number-line estimation. Lack of knowledge of the mappings between numerals and their associated nonverbal mental magnitudes is one of these, as discussed above. Other accounts agree, unsurprisingly, that some form of familiarity with numbers is important for accurate performance, though only the present account makes specific predictions about the potential effects of the acquisition of this knowledge on estimation patterns. Further studies will explore the role of acquiring accurate mappings between numerals and associated nonverbal magnitudes in number-line estimation.

A second source of developmental change lies in the value of the parameter β (the exponent determining the

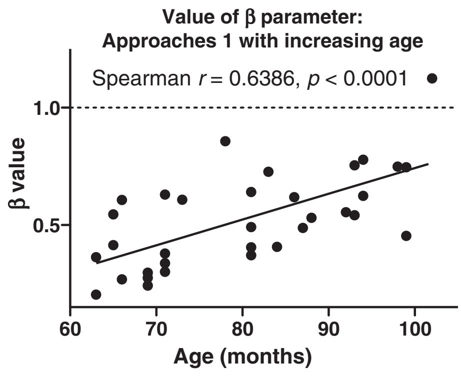

Figure 6 *Values of β parameter vs. age in months. Values of β closer to 1.0 correspond to more accurate (more nearly linear) estimates.*

shape of the power function relating psychological magnitude to actual numerical magnitude). Previous findings with perceptual proportion judgments show that β is closer to 1 in older children (Hollands & Dyre, 2000). This is true in our experiments as well (Figure 6). Because β values near 1 correspond to near-linear relationships, there *is* developmental change toward an increasingly linear representation of numerical magnitude, but there is no evidence of a categorical shift in the type of mental numerical representation used. Rather, in addition to the other sources of change described here, there may be a smooth developmental change in the value of this parameter. Further experiments are necessary to explore this possibility, and to determine the causes of changes in β. The number-line estimation task, like the perceptual proportion tasks it parallels, constitutes a valuable tool for tracking this source of change.

A third potential source of developmental change lies in the tendency to use central reference points. Two children in our 5-year-old group and eight in our 7-year-old group did so (Figures 3 and 5; Tables 1 and 2), although this difference did not reach significance (Fisher's exact test, $p = .067$). Previous researchers have not found age differences in the use of reference points in fourth and sixth graders' proportion judgments with graphical elements (Spence & Krizel, 1994; Hollands & Dyre, 2000), but further studies are needed to explore the contributions of this source of change in numerical proportion judgments.

Conclusions

The proportion-judgment account offers at least five advantages over previous accounts of number-line estimation. First, the proportion-judgment explanation is motivated by the structure of the task: the number-line estimation task is fundamentally a proportion-judgment task. Second, proportion judgment has been modeled and validated in other domains, with many tasks, in children and adults (Spence, 1990; Spence & Krizel, 1994; Hollands & Dyre, 2000; Hollands, Tanaka & Dyre, 2002). Third, this account makes specific, testable predictions about performance patterns: these predictions found support in the present experiments, and the models explained cyclical patterns of estimation bias that remained unexplained by previous accounts. Fourth, the present proportion-judgment account readily explains data produced both by older children, who produce roughly linear-appearing responses for a given number-line task, and by younger children, who produce roughly logarithmic-appearing responses for the same task.

Fifth, this account explains why performance patterns that have led researchers to hypothesize a logarithmic-to-linear shift may arise in number-line tasks, but not in other numerical estimation tasks with similar populations (e.g. estimating and labeling the number of items in a set without the use of number lines: Barth *et al*., 2009; Lipton & Spelke, 2005; Huntley-Fenner, 2001). Number-line tasks can *only* be properly understood as proportion judgments: one cannot place '30' on a number line without knowing that '0' and '100' go at the ends. Other estimation tasks need not be treated as proportion judgments: one can label 30 items as '30' without making a comparison to an anchoring set. Therefore, performance patterns predicted by proportion-judgment models may not appear in all tasks involving numerical estimation. When proportion judgments *are* involved, as they are for number-line tasks, linear-looking or logarithmic-looking estimation patterns do not implicate linear or logarithmic mental representations: the forms of mental representations of number cannot be read so directly from estimation performance.

Acknowledgements

The authors thank the participating childcare centers, parents, and children who made this work possible; Keera Bhandari, Kyle MacDonald, and Jessica Sullivan for assistance with data collection; and Susan Carey and Anna Shusterman for helpful comments. This work was partially supported by a Hughes Summer Research Fellowship to AP.

References

Barth, H., Starr, A., & Sullivan, J. (2009). Children's mappings of large number words to numerosities. *Cognitive Development*, **24**, 248–264.

Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, **41**, 189–201.

Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, **79**, 1016–1031.

Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd edn.). New York: Springer.

Chong, S.C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, **43**, 393–404.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, **320**, 1217–1220.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., & Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year-old children: evidence for a segmented linear model. *Journal of Experimental Child Psychology*, **99**, 1–17.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, **78**, 1343–1359.

Geary, D.C., Hoard, M.K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology*, **33**, 277–299.

Glover, S., & Dixon, P. (2004). Likelihood ratios: a simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin and Review*, **11**, 791–806.

Hollands, J.G., & Dyre, B. (2000). Bias in proportion judgments: the cyclical power model. *Psychological Review*, **107**, 500–524.
Hollands, J.G., Tanaka, T., & Dyre, B. (2002). Understanding bias in proportion production. *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 563–574.
Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': numerical estimation by 5–7-year-olds. *Cognition*, **78**, B27–B40.
Huttenlocher, J., Hedges, L.V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, **98**, 352–376.
Izard, V., & Dehaene, S. (2008). Calibrating the number line. *Cognition*, **106**, 1221–1247.
Laski, E., & Siegler, R. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, **78**, 1723–1743.
Lipton, J., & Spelke, E. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development*, **76**, 978–988.
Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, **103**, 503–515.
Nakajima, Y. (1987). A model of empty duration perception. *Perception*, **16**, 485–520.
Nieder, A., & Miller, E. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, **37**, 149–157.
Noel, M.-P., Rousselle, L., & Mussolin, C. (2005). Magnitude representation in children: its development and dysfunction. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 179–195). New York: Psychology Press.
Opfer, J. (nd). Analyzing the number-line task: a tutorial. Retrieved from http://www.psy.cmu.edu/~siegler/publications-all.html.
Opfer, J., & Siegler, S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, **55**, 169–195.
Shuford, E.H. (1961). Percentage estimation of proportion as a function of element type, exposure time, and task. *Journal of Experimental Psychology*, **61**, 430–436.
Siegler, R.S., & Booth, J. (2004). Development of numerical estimation in young children. *Child Development*, **75**, 428–444.
Siegler, R.S., & Opfer, J. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, **14**, 237–243.
Siegler, R.S., & Ramani, G.B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, **11**, 655–661.
Siegler, R.S., & Ramani, G.B. (in press). Playing linear number board games – but not circular ones – improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology.*
Siegler, R.S., Thompson, C., & Opfer, J. (in press). The logarithmic-to-linear shift: one learning sequence, many tasks, many time scales. *Mind, Brain, and Education*.
Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 683–692.
Spence, I., & Krizel, P. (1994). Children's perception of proportion in graphs. *Child Development*, **65**, 1193–1213.
Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology*, **78**, 392–402.
Varey, C., Mellers, B.A., & Birnbaum, M.H. (1990). Judgments of proportion. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 613–625.

Received: 23 May 2009
Accepted: 4 January 2010

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. *Proportion judgment models used in Experiments 1 and 2*

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.