Developmental Science

Developmental Science 14:5 (2011), pp 1205–1206 DOI: 10.1111/j.1467-7687.2011.01081.x

COMMENTARY

A sense of proportion: commentary on Opfer, Siegler and Young

Hilary Barth,[1] Emily Slusser,[1] Dale Cohen[2] and Annie Paladino[1,3]

1. *Psychology Department, Wesleyan University, USA*
2. *Department of Psychology, University of North Carolina at Wilmington, USA*
3. *WestEd, San Francisco, USA*

This is a commentary on Opfer, Siegler and Young (2011).

A growing body of evidence (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011; Sullivan, Juhasz, Slattery & Barth, 2011) shows that 'linear' estimation patterns in published number-line data actually follow a cyclic power function, a signature of proportional reasoning (Hollands & Dyre, 2000; Spence, 1990). Opfer, Siegler and Young argue that fitting a cyclic power function to number-line data is tantamount to capitalizing on chance. We claim, in contrast, that the cyclic power function is (1) theoretically motivated, (2) mathematically sensitive to important fluctuations in the data, and (3) highly unlikely to be fitting noise in these data. We caution OS&Y that linear functions are insensitive to the systematic fluctuations in the data that describe the observer's estimation bias in the bounded number-line task.

The bulk of OS&Y's substantive critique rests on analyses of previous datasets. But OS&Y's compiled studies were explicitly designed to distinguish between logarithmic and linear models. The observations in their datasets are sparsely distributed near the upper endpoint, where proportion-judgment (PJ) models make specific predictions. Therefore those datasets are biased against PJ hypotheses and toward the logarithmic and linear hypotheses they were designed to capture. We also note that B&P found support for the PJ account in analyses of individuals' estimates, not just group analyses as implied by OS&Y.

The PJ account explains changes in number-line estimation in at least three ways. First, values of Stevens exponents (β) for numerical magnitude might gradually approach 1 with age, leading to more accurate estimates. Second, older children might use more reference points, an independent source of increased accuracy. Third, accuracy requires evaluating the upper endpoint appropriately when deciding where to place a target number (see also Cantlon, Cordes, Libertus & Brannon, 2009). Implicit also in B&P is the possibility that very young children might not recognize the bounded nature of the task at all, instead treating it as an open-ended magnitude judgment; if so their estimates should be well characterized by an unbounded power function.

OS&Y misinterpret the PJ account when they model microgenetic changes in children's estimates as changes in β. They report abrupt changes in β after local feedback, claiming that this finding supports the log-to-linear-shift hypothesis. But in our view, local feedback probably improves accuracy by providing children with a new reference point (predicting broad, not just local, improvements in estimation accuracy; see Hollands & Dyre, 2000, Figure 3). Adoption of a reference point can, of course, occur abruptly. OS&Y unfortunately did not test this hypothesis.

Finally, OS&Y raised the concern that while power models may provide the best *fits*, they are not necessarily the best *predicting* models. They applied leave-one-out cross-validation (LOOCV) techniques to their compiled dataset, finding support for the log-to-linear shift. We conducted LOOCV analyses on B&P's dataset (one better suited to testing alternative hypotheses). We calculated mean squared error (MSE) as the cross-validation error index, instead of using the mean absolute percent error (MAPE) as OS&Y did (MAPE is at times appropriate for evaluating time-dependent series data, but it is not appropriate for datasets discussed here[1]). We also did not separate children with 'logarithmic' vs. 'linear' estimation patterns, a serious problem with OS&Y's analysis (because preassigning these groups

[1] By dividing the error (estimated value minus the actual value) by the actual value, this calculation weighs errors from predictions on the upper end of the scale less than those on the lower end. For example, if an estimated value is 4 and the actual value is 3, the absolute percent error is .33. However, if an estimated value is 91 and the actual value is 90 (again, a discrepancy of only one point), the [M]APE is .01. MSE, on the other hand, weights each of these values equally.

$$MSE = \frac{1}{n}\sum (E - A)^2 \qquad MAPE = \frac{1}{n}\sum \left|\frac{E - A}{A}\right|$$

where E refers to the estimated value and A refers to the actual value.

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

© 2011 Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

essentially circumvents the very safeguards that cross-validation techniques impart, inserting an additional level of model complexity and bias that are not accounted for by these analyses). Our LOOCV results supported the PJ account. For 5-year-olds, an unbounded power model yielded the lowest MSE (26.356), followed by the adapted one-cycle model (27.884) and then the logarithmic model (34.397). The linear, one-cycle, and two-cycle models yielded much larger MSEs (51.278, 66.297, and 210.943, respectively). For 7-year-olds, the two-cycle model yielded the lowest MSE (13.645), followed by the linear model (15.475) and the one-cycle model (19.484; 19.622 for the adapted version), and unbounded power and logarithmic models produced large MSEs (118.908 and 236.527, respectively). Unsurprisingly (see Stone, 1977), the LOOCV analysis describes the models in much the same way as a comparison of AICc scores (Akaike's information criterion, used in B&P's original analyses).

'The choice of a model may be influenced by measures of fit but the final decision concerning which model to use must involve human judgment' (Browne, 2000, p. 110). We agree. Our intent is not to quibble about model selection, but to present theoretical and empirical support for an alternative account of number-line estimation and related paradigms. Given growing evidence that bounded number-line tasks produce the signature ogival functions that indicate proportion estimation, we urge researchers to do the following. First, sample a sufficient number of points along the number line to identify the cyclic function if it is present. Second, model both the cyclic function and linearity to determine relative fits of each model.[2] Limiting analyses to logarithmic and linear models may be sufficient for those who simply need a useful heuristic for classifying children. But for researchers using these tasks to assess the nature of numerical representation, failure to adequately assess and fit proportion models in a bounded number-line task would be a serious omission.

[2] A Microsoft Excel worksheet for performing simple versions of these analyses (Slusser & Barth, n.d.) is freely available at http://hbarth.faculty.wesleyan.edu/publications/.

© 2011 Blackwell Publishing Ltd.

Acknowledgements

This work was supported in part by NSF DRL0950252 to HB and NIH RO1HD047796 to DC.

References

Barth, H., & Paladino, A.M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, **14**, 125–135.

Browne, M. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, **44**, 108–132.

Cantlon, J.F., Cordes, S., Libertus, M.E., & Brannon, E.M. (2009). Comment on 'Log or linear? Distinct intuitions of the number scale in Western and Amazonian Indigene cultures'. *Science*, **323**, 38.

Cohen, D.J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin & Review*, **18**, 331–338.

Hollands, J.G., & Dyre, B.P. (2000). Bias in proportion judgments: the cyclical power model. *Psychological Review*, **107**, 500–524.

Opfer, J., Siegler, R., & Young, C. (2011). The powers of noise-fitting: reply to Barth and Paladino. *Developmental Science*, **14**, 1194–1204.

Slusser, E., & Barth, H. (n.d.) An Excel worksheet for proportion-judgment analyses of number-line data. Retrieved from http://hbarth.faculty.wesleyan.edu/publications/.

Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 683–692.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's Criterion. *Journal of the Royal Statistical Society B*, **39**, 44–47.

Sullivan, J.L., Juhasz, B.J., Slattery, T.J., & Barth, H.C. (2011). Adults' number-line estimation strategies: evidence from eye movements. *Psychonomic Bulletin & Review*, **18**, 557–563.

Received: 18 May 2011
Accepted: 28 May 2011