Cognition 150 (2016) 243–251

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Discussion

No evidence of learning in non-symbolic numerical tasks – A comment on Park and Brannon (2014)

Marcus Lindskog *,[1], Anders Winman [1]

Uppsala University, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 1 October 2014
Revised 27 December 2015
Accepted 7 January 2016

Keywords:
Numerical cognition
Approximate Number System
Cognitive training

ABSTRACT

Two recent studies – one of which was published in this journal – claimed to have found that learning on a non-symbolic arithmetic task improved performance on a symbolic arithmetic task (Park & Brannon, 2013, 2014). This finding has potentially far-reaching implications, because it would constitute evidence for a causal link between the Approximate Number System (ANS) and symbolic-math ability. Here, we argue that, due to the methodology used in both studies, the interpretation of data in terms of an improvement in ANS performance is problematic. We provide arguments and simulations showing that the trends in the data are similar to what one would expect for a non-learning observer. We discuss the implications for the original interpretation in terms of causality between non-symbolic and symbolic arithmetic performance.

1. Introduction

The Approximate Number System (ANS) is thought to be a primitive cognitive system that supports the representation of non-symbolic magnitudes (e.g., Feigenson, Dehaene, & Spelke, 2004). It has been documented in human adults (e.g., Halberda, Ly, Wilmer, Naiman, & Germine, 2012), infants (e.g., Feigenson et al., 2004), and non-human animals (e.g., Brannon, Wusthoff, Gallistel, & Gibbon, 2001). Several studies have indicated that having a more precise ANS is related to better arithmetic ability (e.g., Halberda, Mazzocco, & Feigenson, 2008; Inglis, Attridge, Batchelor, & Gilmore, 2011; Libertus, Feigenson, & Halberda, 2011). This finding has attracted a lot of interest and suggested a causal functional link.

In two recent studies Park and Brannon (henceforth P&B, 2013, 2014) propose that the ANS is causally related to symbolic-math ability. The claim is supported by experimental demonstrations of transfer of learning from a non-symbolic arithmetic, to a symbolic arithmetic task in terms of a math test. P&B (2013, p. 2015) suggested that the results of their study "...show that improvement in an ANS-based, nonsymbolic, approximate-arithmetic training task over multiple sessions transfers to selective improvements in symbolic-math ability." In P&B (2014) the scope was widened by use of several tasks that measured various cognitive components that might be responsible for a causal effect. This strategy aimed at "improving distinct cognitive components" (p. 189) in order to later "compare the transfer effects in exact symbolic arithmetic performance across these training conditions" (p. 189).

These results potentially have very important implications for our understanding of human numerical cognition. For example, Hyde, Khanum, and Spelke (2014, p. 93) argued that the findings of P&B (2013) "...provide the strongest evidence to date of a causal and specialized relationship between the ANS and symbolic mathematics." From an applied perspective, implications are overwhelming. As suggested by P&B (2014, p. 199) the results could mean that approximate arithmetic training could be used in society to "benefit young children who have yet to master the meaning of exact number or numerical symbols".

In both studies, the main conclusions critically depend on the finding that performance improvements on an ANS-based task transferred to a symbolic math task. The logic behind this would be that if X is causally related to Y, an improvement due to training of ability X should induce an alteration of ability Y. More specifically, if it were possible to show that an improvement in ANS-performance by training is accompanied by a subsequent improvement in symbolic arithmetic performance it would suggest a direction of causality from ANS to symbolic math ability. Here, however, we argue that it is not possible to interpret the data provided in the two studies by P&B as showing any improvement in non-symbolic arithmetic at all. We provide simulations suggesting

DOI of original article: http://dx.doi.org/10.1016/j.cognition.2016.02.012
* Corresponding author at: Department of Psychology, Uppsala University, P.O. Box 1225, SE-751 42 Uppsala, Sweden.
E-mail address: marcus.lindskog@psyk.uu.se (M. Lindskog).
[1] M. Lindskog and A. Winman contributed equally to this work.

http://dx.doi.org/10.1016/j.cognition.2016.01.005
0010-0277/

that the trends in data that P&B interpreted as evidence for learning are the trends that one would expect to find for non-learning observers.

2. Park and Brannon's adaptive training method

In the approximate-arithmetic task used by P&B, participants see two arrays of dots moving behind an occluder in sequence (addition task) or one array of dots moving behind an occluder followed by another array of dots appearing from behind the occluder (subtraction task). Participants evaluate the result of the operation (addition or subtraction) implied by the movements of the dot arrays. The responses from participants are elicited in two ways. On comparison trials, participants are presented with a new array of dots and decide if the result of the previously observed operation is more ore less numerous than the new array. On match trials, participants are presented with two new arrays and decide which of the two match the numerosity of the result of the operation. The difficulty of the task is determined by the ratio between the correct answer and the alternative response option.

P&B used an adaptive method, similar to those used in psychophysics (e.g., the "up-down method") to estimate individuals' psychophysical discrimination threshold on various tasks (see e.g., Treutwein, 1995), to train their participants. In their implementation of the method, task difficulty is adjusted after every 20 trials according to how well the participant performed: if performance on the last 20 trials was above 85% correct, the difficulty is increased; if it was below 70% correct, the difficulty is decreased. P&B found that task difficulty stabilized after several sessions, at a level that was considerably harder than the initial one (Fig. 1A, filled dots). Their conclusion from this observation was that performance had improved, indicating learning.[2]

The seemingly very rapid change of stimulus difficulty as a function of training found in all adaptive tasks in both studies (P&B, 2013, 2014) is at first glance striking. From the first to the second training session participants in the different conditions seemingly master visual short-term memory tasks with a higher span, a symbol ordering task at higher speed and an approximate arithmetic task with stimuli much harder to discriminate. Most impressive are maybe the effects found on Approximate Number Comparison, which seem to suggest that performance of participants dramatically improved within 25 min of training with a reduction of weber fractions by two thirds. This finding is surprising considering that other studies (e.g., Lindskog, Winman, & Juslin, 2013) have tried without success to obtain learning by training in very similar tasks (see also DeWind & Brannon, 2012). P&B (2013) suggested that a possible explanation of this discrepancy may lie in the regulation procedure (the adaptive algorithm) "which kept the task challenging", thereby "inducing active engagement" (p. 2017) within participants. While we agree that the regulation procedure embedded in the adaptive algorithm is important per se in understanding the findings, we propose a different explanation than actively engaged participants.

3. "Improvement" without learning

P&B interpreted the increase in difficulty level during training (Fig. 1A, filled dots) as evidence that subjects had gotten better at the task. However, with the adaptive method used by P&B, the direction of convergence (harder/easier) critically depends on the starting value chosen by the experimenter. This intuition is demonstrated in Fig. 1A that illustrates the results of a simulation where participants with equal performance but different starting values take on the task used by P&B. The figure, shows that a relatively easy starting value (filled squares) necessary will lead to convergence on harder stimuli (lower values on the y-axis) whereas a hard starting value (open squares) will bring about convergence on easier stimuli (high values on the y-axis) (see full details about simulations below).

The starting level chosen by P&B was easier than what has been found to be readily mastered by 6-month-old human infants (ratio 1:2) (e.g., Starr, Libertus, & Brannon, 2013; Xu, 2003; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). Hence, it is not surprising that they found that the adaptive method converged to more difficult stimulus levels over time – this is what one would expect, even for observers who do not learn. Had they used a relatively difficult starting level, they would possibly have found – with the same observers – a decrease of difficulty over time. Therefore, the direction of convergence cannot be used to determine whether participants got better at the task.

To establish that learning has taken place, one could instead conduct pre- and post-tests in combination with a proper control group. P&B (2014) did actually obtain pre- and post-tests for three of the measures for which they claim improvement took place with the adaptive tests; Approximate Number Precision, Visuospatial short-term memory, and Numeral order Judgments. Albeit not identical, these tests were very similar to the adaptive tests. Thus, one would expect near transfer effects on these tests if the observed pattern of performance on the adaptive tests were due to learning. No such effects in terms of increased accuracy were found on the pre-post comparisons (an effect was found in terms of faster reaction times on the numerical symbol ordering task).[3] In spite of this finding, the authors interpreted the changes in stimulus difficulty on the adaptive tests in terms of learning.

4. A simulation of a non-learning observer

Another way to establish whether learning has taken place is to compare the human data with data from simulations of a non-learning observer. If the trends in the human data are very different from those predicted for non-learning observers, then the human data may be interpreted as evidence for learning. On the other hand, if simulations of a non-learning observer closely mimic the human data, then it is questionable to conclude that the human data contain evidence for learning.

To investigate what data would look like for a non-learning observer, we performed simulations of Experiments 1 and 2 in P&B (2013) and Experiment 1 in P&B (2014), which we will refer to as E1-2013, E2-2013, and E1-2014, respectively. In these simulations, we used the same procedures as P&B, except that simulating an ideal observer, instead of collecting it from a human observer, generated the response on each trial.

Following previous work (e.g., Barth et al., 2006; Dehaene, 2001; Dehaene & Changeux, 1993) we assume that numerosity estimates are internally represented on a logarithmic scale with constant Gaussian noise. Hence, if we denote the numerosity of a stimulus by N, then the simulated internal representation of this numerosity, n, is drawn from a Gaussian distribution with a mean equal to $\log(N)$ and a standard deviation σ. We further assume that the observers use the optimal decision rule to make their choices.

[2] The present paper focuses on the interpretation of performance in an approximate arithmetic task. However, P&B (2013, 2014) make claims about improvement in performance for tasks involving "numerical ordering", "approximate number comparison", "short-term memory" and "symbol ordering". The main objections presented below of such an interpretation likewise fully apply to all these other tasks.

[3] In order to try to demonstrate near transfer effects, P&B performed post hoc contrasts pooling the approximate arithmetic and the non-symbolic numerical comparison groups. Those analyses approached statistical significance, but did not reach the conventional alpha level of .05.