Journal of Experimental Psychology: General
2014, Vol. 143, No. 1, 255–265

0096-3445/14/$12.00 DOI: 10.1037/a0030825

Hysteresis Affects Approximate Number Discrimination in Young Children

Darko Odic
Johns Hopkins University

Howard Hock
Florida Atlantic University

Justin Halberda
Johns Hopkins University

Perceptual decisions are often affected not only by the evidence gathered during a trial but also by the history of preceding trials. This effect—termed *perceptual hysteresis*—provides evidence for how perceptual information is represented and how it is used. The present research focuses on how the difficulty of preceding trials affects subsequent ones—we find that how well 5-year-old children perform in a 2-alternative forced-choice numerical discrimination task depends on whether they have had a prior history of easier discriminations or a prior history of harder discriminations. Furthermore, this effect is modulated by the feedback children receive. In 3 experiments, we demonstrate that these effects are not related to practice or loss of interest due to negative feedback, or simply to trial difficulty or discriminability. Instead, children appear to have state-dependent confidence states such that prolonged experience making low-confidence decisions degrades performance, whereas prolonged experience making high-confidence decisions improves it. These results are discussed in the context of dynamical psychophysics, representations of confidence, and work on children's and adults' number perception abilities.

Keywords: number discrimination, hysteresis, approximate number system, confidence, dynamical psychophysics

Imagine seeing a set of dots, much like the ones in Figure 1, and being asked to quickly, without counting, identify whether more of the dots are blue or yellow. More than a century of work in adult and developmental psychophysics and visual sciences has used simple perceptual discrimination tasks, much like this one, to establish both how we represent information and how we form decisions about these representations.

Although traditional psychophysics and signal detection theories have largely ignored how perceptual decision making unfolds over time, there have been recent pushes toward a more dynamical theory of psychophysics that describes the gradual integration of evidence and how this accumulation informs later decision making (Busemeyer & Townsend, 1993; Jones, Mozer, & Kinoshita, 2009; Ratcliff & Smith, 2004; Usher & McClelland, 2001; Wagenmakers, 2009). This dynamical approach has been beneficial for extending formal theories to encompass new tasks (Hock, Kogan, & Espinoza, 1997; Trueblood & Busemeyer, 2011; Wagenmakers, 2009), and neurophysiological data (Gold & Shadlen, 2007; Kiani & Shadlen, 2009; Kleinschmidt, Buchel, Hutton, Friston, & Frackowiak, 2002), and has suggested answers to some long-standing questions about aging and development (Ratcliff, Spieler, & McKoon, 2000).

One benefit of dynamical psychophysics is that it can deal with the potentially perplexing effect of perceptual hysteresis. Perceptual hysteresis refers to the commonly observed phenomena that perceptual thresholds (i.e., smallest differences that can be discriminated) change depending on whether stimuli are initially discriminable and gradually adjusted into being nondiscriminable or vice versa (for review, see Hock & Schoner, 2010). For example, Kleinschmidt et al. (2002) asked participants to identify a single letter on a screen that could initially be either very light or very dark, and the letter was gradually made brighter or darker with each trial; the significant contrast at which the letter was identified changed depending on the direction of change—that is, participants were more likely to identify a darker letter when it was gradually changed from light to dark than the reverse. Thus, hysteresis is indicated by the persistence of the initially established percept despite the evidence reaching values that favor the alternative percept.

Traditionally, such hysteresis and order effects have been attributed to artifacts associated with the classical method of limits, including anticipation, inferences about task structure, and response perseverance (Woodworth & Schlosberg, 1972). Recently, however, Hock and Schoner (2010) have demonstrated that, even with all these factors controlled for, perceptual hysteresis is still found in a variety of tasks. Unlike traditional psychophysical and

This article was published Online First November 19, 2012.

Darko Odic, Psychological and Brain Sciences, Johns Hopkins University; Howard Hock, Department of Psychology, Florida Atlantic University; Justin Halberda, Psychological and Brain Sciences, Johns Hopkins University.

This research was funded by NICHD Grant R01 HD057258 awarded to Justin Halberda and an NSERC PGS-D awarded to Darko Odic. Thank you to Anna Zamm and Meg Bowen for help in data collection. We thank Matt Jones for helpful discussions. Portions of this work were presented at the 2011 Vision Sciences Society in Naples, Florida, and at the 2011 Cognitive Science Conference in Boston, Massachusetts.

Correspondence concerning this article should be addressed to Darko Odic, Psychological and Brain Sciences, Johns Hopkins University, Ames Hall, 3400 North Charles Street, Baltimore, MD 21218. E-mail: darko.odic@jhu.edu

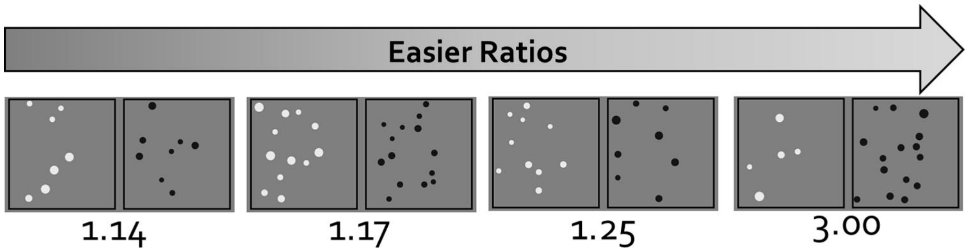

Figure 1. Four examples of stimuli used ranging from harder to easier trials. Each number describes the ratio, calculated as the larger number divided by the smaller number.

signal detection theories, however, dynamical psychophysics does not treat the presence of hysteresis as an unwelcome side effect of testing or participant bias, but rather as evidence for the underlying structure of how perceptual decisions are made. In other words, perceptual hysteresis may be the by-product of how our representations and decision-making mechanisms alter over time in response to making previous decisions (Jones et al., 2009; Kiger & Glass, 1981; Taylor & Lupker, 2001).

For example, in the Hock and Schoner (2010) model, detection of a stimulus depends on a stable network of multiple detectors (rather than single ones, as in classical psychophysics), and continual exposure to similar trials alters the activation thresholds of this network, thus creating a different response to the current trial depending on the previous history. More generally, perceptual hysteresis indicates that a current percept depends in part on the current event and in part on the residual activation state of the neural systems responsible for the immediately preceding percept. Hysteresis effects are induced by preceding events, but what drives the change in the responses are the internal states that the preceding events establish. When these previously established internal states are sufficiently stable they persist long enough to shape perceptual, behavioral, and emotional responses to the current event.

In the present work, we hypothesize that one particular form of hysteresis should depend on the internal states established by the difficulty of previous perceptual decisions. Specifically, most theories of perceptual discrimination argue that any decision is accompanied by one's estimate of confidence in that decision—the subjective likelihood of the decision being correct (for extensive review, see Pleskac & Busemeyer, 2010). Thus, under dynamical psychophysics, the prolonged exposure to easy or difficult trials may result in a high- or low-confidence state that influences future perceptual decisions. One possibility, for example, is that prolonged exposure to low-confidence decisions (i.e., very difficult discriminations) may result in significantly worse performance later on in the task, even when the task becomes easier.

Some previous work suggests that exposure to a block of hard or easy trials alters the decision-making process. Kiger and Glass (1981), for example, gave participants a verification task in which an equation had to be judged as correct or incorrect (e.g., "8 + 10 = 18"). When the identical equation was embedded in a block of difficult equations, response times were significantly higher than when it was embedded in a block of easy equations, irrespective of accuracy. Blocking effects such as these have traditionally been attributed to either a change in decision-making criterion (e.g., threshold for sufficient amount of evidence; Taylor & Lupker, 2001) or to the estimation of the reliability of perceptual evidence (Jones et al., 2009), but have largely been restricted to changes in response times or a speed–accuracy trade-off. Recently, there has also been a suggestion that, on difficult trials, observers guess without even consulting the perceptual evidence (Ludwig & Davies, 2011). Thus, prolonged exposure to difficult trials may also increase one's propensity to give up on the decision and blindly guess, thereby reducing overall performance.

Some evidence also suggests that children's accuracy may be affected by the context of easy or difficult trials. Halberda and Feigenson (2008) tested preschoolers on a simple numerical discrimination task where a display of two groups of items was briefly presented (e.g., Big Bird's and Grover's toys), and children had to identify who had more toys. Children's ability to perform this task depended on the precision of their approximate number system (ANS), a potentially innate mechanism for discriminating numbers of items or events (Dehaene, 2009; Feigenson, Dehaene, & Spelke, 2004; Izard, Sann, Spelke, & Streri, 2009). Critically, discriminations in the ANS obey Weber's law: Discrimination performance depends not on the total number of items in the set but on the ratio between them. In other words, discriminating 10 dots from 9 dots (a ratio of 1.11) is much more difficult than discriminating 10 dots from 5 dots (a ratio of 2.0; see Figure 1). Although it was not the focus of Halberda and Feigenson (2008), many of the children, and especially the youngest ones, showed deviations from Weber's law and appeared to guess on a very high proportion of trials (adults showed much smaller deviations, if at all). Given that the majority of ratios in the task were difficult for children to discriminate, this guessing rate may be related to a hysteresis effect triggered by a series of difficult, low-confidence decisions.

Below, we present three experiments that test for effects of prolonged exposure to difficult or easy trials in 5-year-old children. In particular, we are interested in the possibility that discrimination performance may be affected by confidence-related hysteresis above and beyond the possible effects of trial difficulty and discriminability. Our manipulation in all these experiments is extremely simple—one group of children are presented all of the easiest numerical ratios first and gradually move into the most difficult ones (Easy-First condition); a separate group of children are shown the exact same sequence of trials, except in the reverse order (Hard-First condition). Because feedback is known to focus children on their internal confidence (Bohlmann & Fenson, 2005; Newman & Wick, 1987), we independently manipulated whether children were given accurate feedback (Experiment 1), inverted feedback (i.e., perfectly opposite of correct; Experiment 2), or no feedback (Experiment 3). To foreshadow, we find evidence that the order of trials significantly impacts children's performance in a manner consistent with dynamic signal detection: children are significantly impaired at discriminating easy ratios if they have a history of discriminating difficult ones, and are significantly better at discriminating difficult ratios if they have a history of discriminating easy ones. In Experiment 2 (inverted feedback), we still see the effect of trial order (i.e., Hard-First vs. Easy-First) even though children in the Easy-First condition were given predominately negative feedback, demonstrating that this hysteresis effect is not simply the result of motivation. Finally, in Experiment 3 (no feedback), we find that, when children's decisions are not affected by their confidence states due to the removal of feedback, the manipulation of trial order has no effect.

These results are consistent with the hypothesis that the changes in performance seen in Experiments 1 and 2 are the result of changes in children's high- or low-confidence representations (i.e., a *confidence hysteresis* effect). This effect, in turn, has implications for children's learning and suggests that properly scaffolding children by giving them the easiest trials first may result in more robust discrimination performance (cf. Siegler & Jenkins, 1989; Vygotsky, 1978).

Experiment 1

Method

Participants. Twenty children (10 per condition) between the ages of 4 and 5 years participated; all were recruited by phone and e-mail from a database maintained in the Lab for Child Development at Johns Hopkins University. The average age of the children was 4.93 years ($SE = 0.16$). All were recruited from the local Baltimore community and were generally middle-class. An additional child was run but excluded from the study for fussiness and not completing the task.

Methods. Stimuli were blue and yellow dots presented on a Macbook Pro laptop with a 13" screen with custom-made Java programs to display the stimuli. The screen was divided into two sections by two rectangular boxes (see Figure 1). Yellow dots always appeared in the left box and blue dots in the right box. On each trial, participants attempted to determine which set (i.e., left or right) had the greater number of dots, with correct side being random across trials. To discourage the use of nonnumeric cues, such as the size of the dots, the displays were controlled for the total size of the individual elements: On half of the trials the larger set by number also had larger total area (Congruent trials), and on the other half the smaller set by number had a larger total area (Incongruent trials). Numerical ratio for each trial was calculated by dividing the larger number of dots by the smaller; the ratios used were 1.11 (e.g., 10 blue: 9 yellow dots), 1.14 (8:7), 1.17 (14:12) 1.25 (10:8), 1.50 (9:6), 2.00 (10:5), and 3.00 (15:5). Each ratio was presented five times, yielding a total of 35 trials.

The complete trial-by-trial sequence of ratios in each condition is presented in the Appendix. For the Easy-First condition, we created a pseudorandom order of trials that began with the easiest ratios (e.g., 3.00, 2.00, and 1.50) and gradually moved into the most difficult ones (e.g., 1.17, 1.14, and 1.11). This trial sequence was exactly reversed in the Hard-First condition. All aspects of the task and displays (e.g., spatial position of individual dots and side of presentation) were identical between the two conditions and the only difference was the trial order.

Procedures. Each participant was tested individually in a room with the experimenter, and the test session was digitally recorded. The participant sat approximately 60 cm from the computer screen. Parents were seated in the corner of the room and could not see the stimuli, thereby preventing any inadvertent cuing. Children were told that they could play a game where they would see some blue and yellow dots and would have to answer whether "More of the dots are blue or yellow" on each flash. The children were also told that "This game will sometimes be easy, and sometimes will be hard." During six practice trials the blue and yellow dots appeared separately for 1,200 ms each; the six practice trials contained only the three easiest ratios, regardless of whether the practice preceded Hard-First or Easy-First orders. During the actual trials, both sets of dots appeared simultaneously and stayed on the screen for 1,200 ms. Children could respond either by saying the color or by pointing to the screen; the experimenter would push a button on the laptop to record response time (RT) and the child's answer. Participants received accurate feedback after each trial in the form of a computerized voice saying: "That's right!" or "Oh no, that's not right" depending on their answer (see also Halberda & Feigenson, 2008). Once finished, the children were congratulated on doing well on the task and were taken back to the waiting room for a diploma and a prize.

Results

We first verified that performance did not differ as a function of area-congruent versus area-incongruent displays as revealed by a 2 (condition: Easy-First, Hard-First) × 2 (size: congruent, incongruent) mixed-measure measures analysis of variance (ANOVA) that yielded neither a significant effect of size, $F(1, 18) < 1$; $\eta_\rho^2 = 0.008$, nor a Size × Condition interaction, $F(1, 18) < 1$; $\eta_\rho^2 = 0.02$. This suggests that children were using numeric cues in both conditions. As a result, area congruent and incongruent trials were combined and analyzed together throughout.

The average percent correct across the two conditions and seven ratios is presented in Figure 2. These data were analyzed via a 2 (condition: Easy-First, Hard-First) × 7 (ratio) mixed- ANOVA. We found a significant main effect of ratio, $F(6, 108) = 10.26$; $p < .001$; $\eta_\rho^2 = 0.36$, with children performing better with easier ratios. We also found a significant main effect of condition, $F(1, 18) = 9.35$; $p < .01$; $\eta_\rho^2 = 0.34$, with better performance on the Easy-First condition overall at a mean of 74.57% ($SE = 3.50$) and worse performance on the Hard-First condition at 59.43% ($SE = 3.50$). As can be seen in Figure 2, the Hard-First children failed to reliably discriminate even the easy ratio of 1.5, even though, on the basis of the performance of the children in the Easy-First condition, they clearly had the ability to do so. There was also no significant Ratio × Condition interaction, $F(6, 108) = 1.35$; $p = .24$; $\eta_\rho^2 = 0.07$, although the majority of the difference between the Easy-First and Hard-First conditions resides in the easier rather than harder ratios (Figure 2).

We also examined RTs. To verify that this correctly reflected children's actual RTs, we coded RTs offline from the recorded videos of each testing session by measuring the time between the dots appearing and the child saying his or her response. For each child, individual trial RTs were removed if they fell above or

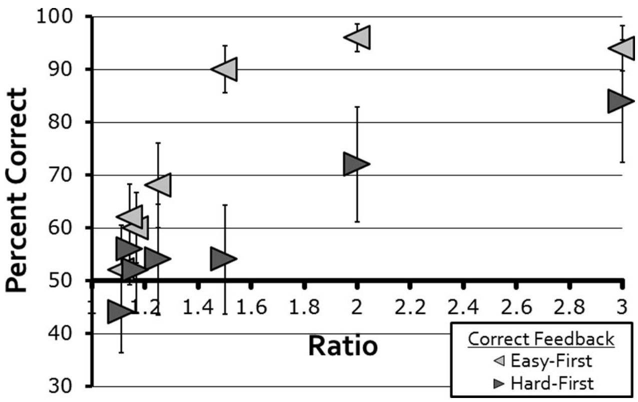

Figure 2. Average percentage correct across seven ratios and the two conditions of Experiment 1. Bars represent standard error of the mean. Symbols indicate the direction of trial order (e.g., Hard-First children would have experienced trials generally decreasing in difficulty reading from left to right on the *x*-axis over the course of the task).

below 2.5 standard deviations from that child's mean (3% of total trials). We then computed a 2 (condition) × 7 (ratio) mixed-measures ANOVA over these video-coded RTs. As expected, there was a significant main effect of ratio, $F(6, 108) = 3.14$; $p < .001$; $\eta^2_\rho = 0.15$, with RTs getting shorter with easier ratios. There was also a nonsignificant trend of condition, $F(1, 18) = 3.14$; $p = .09$; $\eta^2_\rho = 0.149$, with children in the Easy-First condition making their numerical decisions *faster* ($M = 2{,}283.45$ ms; $SE = 143.44$ ms) than the children in the Hard-First children ($M = 3{,}060.65$ ms; $SE = 413.08$ ms). Thus, children in the Easy-First condition were both more accurate and marginally faster than children in the Hard-First condition—both measures indicating superior performance in the Easy-First condition. There was also no significant Ratio × Condition interaction, $F(6, 108) = 1.335$; $p = .25$; $\eta^2_\rho = 0.069$.

To estimate whether condition affected children's discrimination sensitivity, we psychophysically modeled the performance of the children in accordance with Weber's law. Such a procedure is common in number discrimination tasks (Halberda & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011; Meck & Church, 1983; Piazza et al., 2010; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004) and allows us to estimate the Weber fraction—that is, an estimate of internal noise and discrimination sensitivity. The Weber fraction (or *w*) describes the function that predicts percentage correct at every possible ratio.

To determine *w*, we applied a psychophysical model—Equation 1—previously used in Halberda, Mazzocco, and Feigenson (2008), Halberda and Feigenson (2008), Libertus et al. (2011), and Piazza et al. (2010):

$$percent\ correct = \frac{1}{2}\,erfc\left(\frac{n_1 - n_2}{\sqrt{2w}\sqrt{n_1^2 + n_1^2}}\right) * 100 \qquad (1)$$

The model assumes that the underlying representations of number are distributed along a continuum of Gaussian random variables. The equation models the representations for the numerosities presented on each trial as Gaussian random variables, subject to the internal noise of the system (i.e., *w*)—one having a mean of n_1, and the other with a mean of n_2. An important implication of this model is that the two numerosities on each trial will have overlapping representations. As the two quantities become increasingly closer to one another (i.e., closer to a ratio of 1.0), their representations overlap more and participants have a more difficult time determining which is larger, resulting in decreasing accuracy at the task as a function of ratio. The model uses the complementary error function *erfc* to estimate the expected percentage correct at each possible ratio, producing a smooth function that can be compared to the actual observed data.

This model has only a single free parameter—the Weber fraction (*w*)—which indicates the amount of noise in the underlying Gaussian representations (i.e., the standard deviation of the n_1 and n_2 Gaussian representations such that $SD_{n1} = w \times n_1$). Larger *w* values indicate higher representational noise and, thus, poorer discrimination sensitivity across ratios (lower Weber fractions indicate better performance). For each condition, the *w* value that minimized the least squared error was selected as the best fitting one.

Because each child completed only 35 trials, variability was such that we could not reliably fit each child's data to find their *w* value. Instead, the children's data in each condition were grouped together for the Easy-First and Hard-First conditions, and these combined data were used to find the *w* value for each condition. The *w* value that resulted in the best fit of the model to the children's percentage correct for the Easy-First condition was 0.29 ($r^2 = 0.94$), which is lower (i.e., better) than previous estimates for 5-year-olds (e.g., Halberda & Feigenson, 2008), suggesting that these children did well on the numerical discrimination task. The best fit value for the Hard-First condition was 0.86 ($r^2 = 0.89$); a *w* this high is typically associated with the performance of 9-month-old infants (Xu & Spelke, 2000), further supporting the point that these children did extremely poorly.

Discussion

In our first experiment, children were given a simple number discrimination task and had to judge whether more of the dots were blue or yellow on each trial. Unknown to them, we manipulated the order of trials and gave them either the easiest ratios first or the hardest ratios first. We found a large effect of condition: Children in the Hard-First condition performed extremely poorly, with a Weber fraction (*w*) resembling that of 9-month-old infants (Xu & Spelke, 2000). Their performance was poor even at the easiest ratios, suggesting that the history of difficult, low-confidence decisions reduced performance later on. In contrast, children who saw easier trials first performed well above chance, even on a difficult ratio of 1.17 (14:12 dots).

Although the effect appears robust, it is not yet clear what is driving the difference between these two conditions. One possibility is that children are affected not by the difficulty of the trials and their internal confidence, but by the feedback itself (e.g., "That's right!" or "Oh no, that's not right" on each trial). Because children in the Hard-First condition experienced more difficult ratios early on, they also received more negative feedback early on, and may have felt demotivated as a result (for review on feedback's influence on motivation, see Balzer, Doherty, & O'Connor, 1989; Kluger & DeNisi, 1996; Vroom, 1964). Thus, it may be that the observed hysteresis effect emerges whenever the participant is given consistently negative feedback, regardless of the actual difficulty of the trials and the participant's internal confidence in his or her judgments.

To investigate the impact of negative feedback on motivation to do the task, in Experiment 2 we instructed a new group of children to determine which side had more dots but also gave them *inverted* feedback during the task—that is, the computer told them they were right when they were wrong, and vice versa. Thus, children in the Easy-First condition of Experiment 2 got predominantly negative feedback early on, while children in the Hard-First condition got a mix of positive and negative feedback early on. Our interest was *not* whether children in the Easy-First condition would use the inverted feedback to learn to respond oppositely to the stated rule, but rather that hearing "Oh no, that's not right," on most of the trials would lead them to lose motivation and perform poorly. If the observed hysteresis effect emerges entirely due to hearing negative feedback, we would expect results to be the opposite of those in Experiment 1; children in the Easy-First condition should perform near chance levels whereas children in the Hard-First condition should do better.

Experiment 2

Method

Participants. Twenty children (10 per condition) participated in this experiment; none had participated in the previous experiment. The average age was 5.00 years ($SE = 0.16$). Three additional children were tested but were removed from the analysis; two of them (one in each condition) realized that the feedback was inverted and exclaimed to the experimenter that they would start giving opposite answers to match the feedback. One additional child did not complete the task due to fussiness. Results were unchanged if these children were retained.

The two excluded children are worth noting. These two children also performed significantly below chance on the task (i.e., they systematically chose the set that was fewer in number). All other children in both conditions performed above chance on the task throughout. However, it is still possible that some children did realize that the feedback was inverted but chose to ignore it. One's ability to rely on internal confidence in order to disregard feedback is an interesting topic in its own right, but our focus here is solely on whether negative feedback alone can produce the effects observed in the first experiment.

Methods and procedures. To make sure that children understood the task, the practice trials were identical to the first experiment's, and children received valid feedback. During the test trials, this feedback was exactly the opposite from the child's actual performance. Thus, if the child got the answer right according to the numerically-more rule, the computer informed them that they got it wrong, and vice versa. Parents were fully informed of this deception and explicitly agreed to have their child participate. In order to debrief the children, after the experiment was done, children were asked if the game they played was a "more dots" or "less dots" game; afterward, they were told that the computer was being tricky, that the feedback was not accurate and that they did a great job at the task. All other aspects of the experiment (e.g., trial order and displays) were identical to Experiment 1.

Results

The average percentage correct across the two conditions and seven ratios is presented in Figure 3. These data were analyzed by a 2 (condition: Easy-First, Hard-First) × 7 (ratio) mixed-measures ANOVA. We found a significant main effect of ratio, $F(6, 108) = 8.41$; $p < .001$; $\eta_p^2 = 0.32$, with children performing better with easier ratios. We also found a significant effect of condition, $F(1, 18) = 6.01$; $p < .05$; $\eta_p^2 = 0.25$, with children in the Easy-First condition being more accurate at determining the numerically-more set (average accuracy = 69.71%, $SE = 3.45$) compared to children in the Hard-First condition (average accuracy = 57.71%, $SE = 3.45$). Additionally, we found a significant Condition × Ratio interaction, $F(6, 108) = 2.41$; $p < .05$; $\eta_p^2 = 0.12$ (see Figure 3). These results are consistent with the hysteresis effect demonstrated in Experiment 1. Furthermore, and in parallel with the first experiment, we failed to find any effect of condition on RT, $F(1, 18) < 1$; $p = .77$.

As in Experiment 1, we modeled the group performance from each condition according to Weber's law using Equation 1. The psychophysically fit w value for the Easy-First condition was 0.41 ($r^2 = 0.82$); we could not adequately fit a w value to the Hard-First condition, as their performance did not sufficiently improve with ratio. The lowest mean squares value occurred at $w = 1.15$ (with the extremely low $r^2 = 0.06$), a w value worse than that typically found in 6-month-old infants (Halberda & Feigenson, 2008; Xu & Spelke, 2000).

These results are consistent with the pattern observed in Experiment 1 and robustly demonstrate that inverting feedback did not reverse the hysteresis effect—as would be observed if this effect derived solely from children becoming discouraged at hearing a preponderance of negative feedback early in the trials. That is, hearing mostly negative feedback early in the experiment did not hurt the performance of children in the Easy-First condition, and hearing a mix of positive and negative feedback early in the experiment did not help the performance of children in the Hard-First condition.

To confirm that these results were not significantly different from the first experiment, we compared the performance of the participants in Experiments 1 and 2 via a 2 (experiment: correct feedback, invalid feedback) × 2 (condition: Easy-First, Hard-First) × 7 (ratio) mixed-measures ANOVA. There was no main effect of experiment, $F(1, 36) < 1$, nor was there a significant

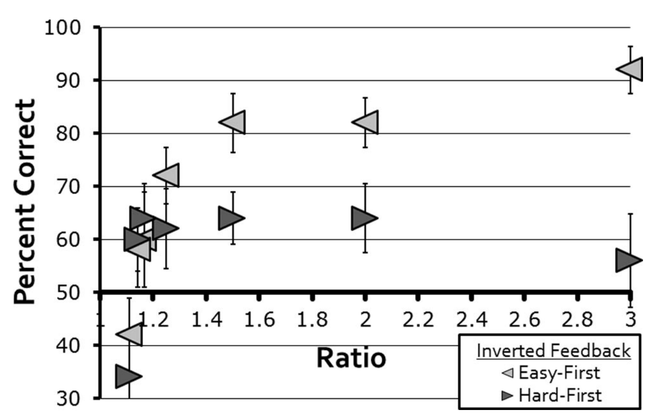

Figure 3. Average percentage correct across seven ratios and the two conditions of Experiment 2. Bars represent standard error of the mean.

Experiment × Condition interaction, $F(1, 36) < 1$, whereas the differences between the Easy-First and Hard-First conditions remained significant: effect of condition, $F(1, 36) = 15.25$; $p < .01$; $\eta_p^2 = 0.30$ and ratio, $F(6, 216) = 17.03$; $p < .01$; $\eta_p^2 = 0.32$. Thus, children performed comparably in these two experiments.

Discussion

In Experiment 2, we asked if feedback alone could be responsible for the difference between Hard-First and Easy-First conditions. To that end, we gave children in the Easy-First condition predominantly negative feedback; however, children in this condition did well and significantly outperformed the Hard-First condition. Note that, ironically, because they continued to do well, the Easy-First condition children got much more negative feedback than the Hard-First condition children in this experiment. These results suggest that the difference in the amount of negative or positive feedback could not alone be responsible for the hysteresis effect found in the first experiment.[1]

These results, however, do not yet conclusively point to the source of the hysteresis effect. For example, it is not clear whether the effect emerges because of changes in trial difficulty or because of changes in internal confidence. Although difficulty and confidence are robustly correlated, they are not identical (Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2012; Wilimzig, Tsuchiya, Fahle, Einhäuser, & Koch, 2008). As such, our data are entirely consistent with a form of hysteresis that is driven by the experienced difficulty of previous trials, or by a form of hysteresis that is driven by a series of low- or high-confidence decisions on previous trials.

Separating confidence and difficulty is an especially daunting challenge, given the tight relationship between the two (Pleskac & Busemeyer, 2010). Previous work, however, has demonstrated that, in the absence of feedback, children do not robustly or reliably report their confidence states. Newman and Wick (1987) demonstrated that children as old as 12 years require the presence of feedback in order to accurately report their confidence in a number estimation task. Similarly, Bohlmann and Fenson (2005) showed that 3- to 5-year-olds accurately monitor and adjust their performance in a card-sorting task only when feedback is provided. Thus, one method of separating difficulty from confidence would be to remove feedback entirely.

In order to establish whether the hysteresis effects observed in the first two experiments are driven by trial difficulty or by internal confidence, Experiment 3 tested children on the identical task but in the absence of any feedback. If children continue to show differences between Easy-First and Hard-First conditions in the absence of feedback, then trial difficulty, which should not be impacted by the absence of feedback, is likely driving the hysteresis effect. If, on the other hand, we no longer find an effect of condition, then changes in internal confidence are the most likely source of the hysteresis effect demonstrated here.

Experiment 3

Method

Participants. Twenty children (10 per condition) participated in this experiment; none had participated in the previous experiments. The average age was 4.95 years ($SE = 0.12$). One child was removed from the analysis for fussiness and not completing the task.

Methods and procedures. This experiment was identical to the first in every way except that, once the practice trials were over, children received no feedback from the computer or the experimenter. Feedback remained on during the six practice trials in order to make sure that children understood the task.

Results

The average percentage correct across the two conditions and seven ratios is presented in Figure 4. These data were analyzed via a 2 (condition: Easy-First, Hard-First) × 7 (ratio) mixed-measures ANOVA. We found a main effect of ratio, $F(6, 108) = 9.39$; $p < .001$; $\eta_p^2 = 0.34$, with children performing better with easier ratios. Critically, we found no significant effect of condition, $F(1, 18) < 1$; $p = .67$; $\eta_p^2 = 0.01$, nor did we find a significant Ratio × Condition interaction, $F(6, 108) < 1$; $p = .51$; $\eta_p^2 = 0.05$. We also found no effect of condition on RTs, $F(1, 18) = 2.22$; $p = .15$.

When psychophysically modeled using Equation 1, we found the Easy-First no feedback condition w to be 0.53 ($r^2 = 0.91$) and the Hard-First no feedback w to be 0.52 ($r^2 = 0.72$)—that is, nearly identical w values.

We next compared whether children in Experiment 3 performed differently from children in Experiment 1. A 2 (experiment: correct feedback, no feedback) × 2 (condition: Easy-First, Hard-First) × 7 (ratio) mixed-measures ANOVA yielded a significant Experiment × Condition interaction, $F(1, 36) = 5.18$; $p < .05$; $\eta_p^2 = 0.13$, with the post hoc contrast revealing that the Easy-First correct feedback children performed significantly better than the Easy-First no feedback children, $t(18) = 2.57$; $p < .05$. These results suggest a significant role for the *presence* of feedback in driving the confidence hysteresis effects observed in Experiment 1. As discussed below, previous work has suggested that the presence of feedback is the necessary contextual cue for children's performance to be affected by their confidence (Bohlmann & Fenson, 2005; Newman & Wick, 1987). The presence of feedback in the first two experiments likely allowed children to, for better or for worse, alter their decisions in response to their confidence states, thereby resulting in the observed hysteresis effect. Alternatively, the lack of feedback did not allow children to form stable confidence states, thereby removing the hysteresis effect.

General Discussion

In the current study we used a number discrimination paradigm to determine whether the history of preceding discriminations

[1] A recent finding from Shibata, Yamagishi, Ishii, and Kawato (2009) found that adults in an inverted feedback task perform better than those in an accurate or no feedback task. However, the difference between their findings and ours may be both in the population (adults vs. children) and in the task. Their subjects received feedback at the end of each block, and the feedback was in the form of total percentage correct; their subjects, therefore, did not know on a trial-by-trial basis what their performance was like and could more gradually adjust their performance. Additionally, it is possible that children and adults react differently to feedback (O'Leary & O'Leary, 1977).

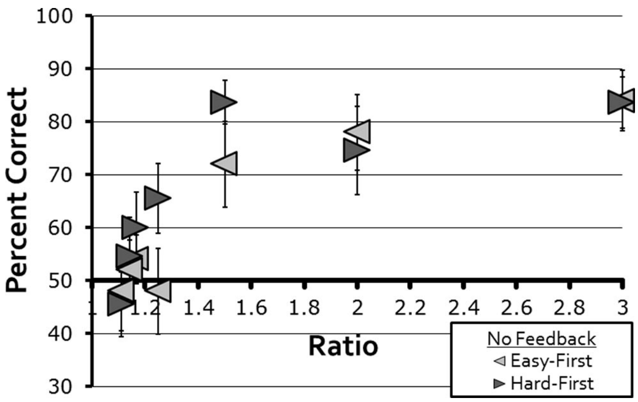

Figure 4. Average percentage correct across seven ratios and the two conditions of Experiment 3. Bars represent standard error of the mean.

impacts decision making on later trials. In three experiments we demonstrated that, in a simple numerical discrimination task, performance on any given trial may drastically change depending on whether the previous history of numerical decisions was made with high- or low-confidence. In other words, we have identified an effect of *confidence hysteresis*—recurrent low-confidence decisions result in significantly degraded performance later on in the task, even on easy trials, and recurrent high-confidence decisions result in significantly improved performance later on, even on difficult trials. These results support dynamical psychophysics and suggest that an observer's discrimination ability depends not only on the characteristics of the external stimuli, but also on the characteristics of the perceptual system that can be altered through the history of previous trials.

Our three experiments demonstrate that certain explanations of confidence hysteresis, including changes in the speed–accuracy trade-off, practice effects, or demotivation due to negative feedback, cannot be solely responsible. For example, it is possible that children in the Hard-First condition—distracted by receiving a barrage of difficult trials early in the task—might have forgotten what the decision rule was, while children in the Easy-First condition received repeated reinforcement of this rule. However, the absence of a hysteresis effect in Experiment 3 is telling against this possibility, as children in both the Hard-First and Easy-First conditions of Experiment 3 retained the correct rule throughout the task. Another explanation of our hysteresis effect may be that negative feedback alone could produce the effect by changing children's motivation to do the task. However, Easy-First children in Experiment 2 remained resolute in the face of predominantly negative feedback, and Hard-First children in Experiment 2 were not aided by hearing a different mix of positive and negative feedback.

Because we did not directly measure confidence, another concern is that the hysteresis effect demonstrated here derives from trial difficulty rather than internal confidence. However, Experiment 3 maintained the identical trial order, trial difficulty, and discriminability as Experiments 1 and 2, and yet we failed to find a hysteresis effect. As reviewed above, the critical manipulation in Experiment 3—the absence of feedback—has previously been shown to prevent children from forming stable confidence states and reduces their ability to adjust their behavior in response to internal confidence (Bohlmann & Fenson, 2005; Newman & Wick, 1987). In contrast, we know of no studies that have shown that, in the absence of feedback, stimulus discriminability or trial difficulty will change. This suggests that the hysteresis effect demonstrated here derives primarily from changes in internal confidence states. This conclusion is also consistent with the dynamic psychophysics, wherein changes in the states of the perceptual system (e.g., confidence), and not just characteristics of the external stimuli (e.g., discriminability), affect the decision making of the observer.

What is the mechanism by which a series of low- or high-confidence decisions may affect children's performance on later trials—that is, how does confidence hysteresis result in better or worse discrimination performance? The first possibility, which is consistent with the very minor improvement in performance with ratio for children in the Hard-First condition of Experiment 2, is that a series of low-confidence decisions might lead a participant to simply give up on the task and guess randomly on a trial. Such guessing without consulting the perceptual evidence has recently been suggested to play a central role in dynamic adult decision making (Ludwig & Davies, 2011). This guessing hypothesis is also consistent with the relatively flat increase observed for children in the Hard-First condition between the hardest and medium-difficulty ratios in Experiment 1 (Figure 2), followed by the sudden improvement in performance when the trials became extremely easy. That is, children may have given up on trying to discriminate the sets until the trials became so easy as to inspire them to re-attend to the task. Although the fitted model of Equation 1 predicts a gradual improvement over ratios, a more parsimonious model may be one where there is flat performance up to a point, and then, once the trials become easy enough, a sudden increase in performance with ratio. This pattern is also observed in traditional perceptual hysteresis, where observers very suddenly switch between two stable perceptual states (Hock & Schoner, 2010). A simple sigmoidal function was previously proposed by Halberda and Feigenson (2008) to fit this kind of behavior, but specification of a psychologically plausible model remains an area for future work. Finally, the guessing hypothesis suggests that the superior performance of the Easy-First children is due to their lack of any guessing and using the perceptual information on every trial, even when ratios became difficult.

A second possibility is that our effects are due to children learning *which* representations to use or not to use in the task. For even a simple discrimination, there are often multiple sources of evidence (e.g., number, cumulative area, density) as well as multiple procedures one might use to generate an answer (e.g., discrete sampling, weighted combination; for review, see Goldstone, 1998). As such, feedback throughout the task in combination with an internal sense of high or low confidence might serve to direct a participant toward the perceptual dimensions and decision rules that seem to be most informative given the context. Consistent with this suggestion, adults are occasionally seen to rely on cumulative area to make a numerical discrimination (Hurewitz, Gelman, & Schnitzer, 2006; Tokita & Ishiguchi, 2010) or to use line length instead of number where displays contain dots that are properly arranged into lines (Pietroski, Lidz, Hunter, & Halberda, 2009). Likewise, children who are not good at counting have been shown to rely on approximate estimates of number, despite having just counted the numbers exactly (Michie, 1984). Under this explanation, low confidence is a signal that an alternative discrimination strategy should be adopted: for example, adult observers appear to use cumulative area for making a numerical discrimination only when trial difficulty is such that using approximate number representations results in difficult, low-confidence decisions (Hurewitz et al., 2006). Similarly, children in the Easy-First conditions of Experiments 1 and 2 may have been properly scaffolded and appropriately relied on the high-confidence signals of their ANS number representations throughout the task. In contrast, the Hard-First children of Experiments 1 and 2 were faced with the difficult challenge of maintaining their attention on the relevant numerical dimension in the face of repeated difficult and low-confidence numerical decisions. Under this interpretation, the mixed positive and negative feedback signals of Experiments 1 and 2 may have served to inform these children that they should search for some other source of evidence for making their decision (e.g., cumulative area, density, contour length). These children either responded randomly, not knowing what source of evidence to use, or constantly used different sources of evidence, none of which produced reliable judgments.

At the present moment, our data cannot adjudicate between these two explanations for the confidence hysteresis effect demonstrated here (i.e., give-up or learning), and future studies, including those testing whether confidence hysteresis transfers to other tasks, will be required. Indeed, it seems likely that internal confidence states are relevant for both deciding when to give up on a task and deciding how to make adjustments within a task.

We also observed a role for the *presence* of feedback. We suspect that the feedback had two effects on the children. First, anecdotally, children in the first two experiments (i.e., with feedback) seemed much more engaged in the task; feedback may act as a motivator for children to try to attend more and try harder on the task. Second, consistent with previous research (Bohlmann & Fenson, 2005; Newman & Wick, 1987), feedback was likely the relevant trigger for children to adjust their behavior in response to their low- or high-confidence states, or to form these confidence states in the first place. In turn, a series of low- or high-confidence decisions may have resulted in compensatory strategies such as those discussed above (e.g., being demotivated and guessing or learning to switch to some other source of evidence). The results of Experiment 3 suggest that, in the absence of adjusting behavior to one's internal confidence, decision making strategies may go unchanged and the confidence hysteresis effect may be avoided. This result is consistent with several models of feedback's influence on motivation and learning (Balzer et al., 1989; Kluger & DeNisi, 1996; Vroom, 1964).

Our findings have important consequences for several domains. First, they confirm that perceptual decision making is not just the product of momentary representations, but that the history of previous decisions matters for the one at hand (see also Kiger & Glass, 1981). In other words, our findings demonstrate that dynamical psychophysics and signal detection theories, in which current decisions are affected not only by the characteristics of the external stimulus but also by the history of previous decisions, are more likely to correctly capture behavior in discrimination tasks. Additionally, our findings demonstrate that internal confidence—which has a long history in psychophysics (though often a secondary role to accuracy and RT)—has serious repercussions on observed performance. In the present experiments we find this especially interesting, because at no point were the children asked to rate their confidence; in other words, our task indirectly manipulated confidence to alter the accuracy of any given decision.

Confidence hysteresis also has implications for children's learning. The effects observed here could be considered a form of scaffolding, whereby the ease and high-confidence of the initial trials allowed children to learn more quickly, or perhaps to remain more motivated, for later trials. Scaffolding has been an important factor in improving children's performance in many other domains, including categorization (Kotovsky & Gentner, 1996), math problem solving (Siegler & Jenkins, 1989), and cognitive control (Brace, Morton, & Munakata, 2006). Recently, a series of studies by Weinstein and Roediger (2010, 2012) also demonstrated that a Hard-First order of trials reduces the subsequent confidence and evaluation of test performance in adults (though their actual performance was unimpaired). The effect reported here supplements these findings and suggests both a potential mechanism by which children could perform better on a variety of scaffolding tasks and that scaffolding may be beneficial to any task in which children internally represent confidence.

Our findings also have important implications for cognitive development and the study of number approximation. There has been a recent surge of experiments on the hardest ratio that children can reliably discriminate (i.e., w). First, w has been shown to gradually improve over the lifespan, although the causes of this change remain unknown (Halberda & Feigenson, 2008; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Odic, Libertus, Feigenson, & Halberda, 2012). Second, lower w values (i.e., better discrimination performance) are related to better performance on math tests in preschoolers (Libertus et al., 2011), school-aged children (Gilmore, McCarthy, & Spelke, 2010), adolescents (Halberda, Mazzocco, & Feigenson, 2008), and adults (Libertus, Odic, & Halberda, in press; Lyons & Beilock, 2011) across the lifespan (Halberda et al., 2012). Third, individuals with math learning disabilities have significantly worse w scores (Mazzocco, Feigenson, & Halberda, 2011; Piazza et al., 2010). Our findings demonstrate that one can capture very different w values depending on the order of trials—that is, children who tend to get the hardest trials first will end up having worse w values than those who get the easiest trials. Although our findings, by directly manipulating the order, have probably magnified this effect, it is entirely possible that simple randomization of trials in a regular discrimination task will, for some participants, result in predominantly difficult trials being first. In fact, when looking at Halberda and Feigenson's (2008) findings, it is clear that many of the younger children show patterns very similar to our data—they do not appear to show a clear improvement in ratio until the easiest ratios. Given that the parameters of the task in Halberda and Feigenson (2008) included ratios that are, for the most part, difficult for children of this age, it is possible that they were underestimating the w values for children in this age range (though note that the sigmoidal fits also presented in that article would partially address this concern). We also found that children perform significantly better with feedback compared to without feedback. Thus, one potential solution for future studies seeking to measure w values would be to administer the easiest trials first, with feedback, and gradually move into the most difficult ones. Our results suggest that this would determine an upper bound for each child's best possible performance. Other authors may be interested in a lower bound as might be assessed by running trials in a Hard-First order with feedback.

Finally, although we focused our investigation on numerical discrimination in children, there is no reason to believe that only this population and only this task will be subject to confidence hysteresis effects. In fact, given dynamical psychophysics, virtually any perceptual decision could be affected by the confidence one has experienced during prior decisions. This is consistent with the blocking effect literature, which has found that easy or hard blocks impact the RT on, for example, sentence verification tasks, equation verification tasks, and word naming latencies (see Jones et al., 2009). The generality of confidence hysteresis, which may motivate an observer to reduce attention to information and discriminations that have resulted in low confidence in the past, may be adaptive and beneficial for future decision making by motivating observers to exploit other sources of evidence and to attempt other types of discriminations that may be of greater value or generate greater confidence.

References

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106*(3), 410–433. doi:10.1037/0033-2909.106.3.410

Bohlmann, N. L., & Fenson, L. (2005). The effects of feedback on perseverative errors in preschool aged children. *Journal of Cognition and Development, 6*(1), 119–131. doi:10.1207/s15327647jcd0601_7

Brace, J. J., Morton, J. B., & Munakata, Y. (2006). When actions speak louder than words: Improving children's flexibility in a card-sorting task. *Psychological Science, 17*(8), 665–669. doi:10.1111/j.1467-9280.2006.01763.x

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432–459. doi:10.1037/0033-295X.100.3.432

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences, 1156*(1), 232–259. doi:10.1111/j.1749-6632.2009.04469.x

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. doi:10.1016/j.tics.2004.05.002

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394–406. doi:10.1016/j.cognition.2010.02.002

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612. doi:10.1146/annurev.psych.49.1.585

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. doi:10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *PNAS: Proceedings of the National Academy of Sciences, 109*(28), 11116–11120. doi:10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668. doi:10.1038/nature07246

Hock, H. S., Kogan, K., & Espinoza, J. K. (1997). Dynamic, state-dependent thresholds for the perception of single-element apparent motion: Bistability from local cooperativity. *Attention, Perception, & Psychophysics, 59*(7), 1077–1088. doi:10.3758/BF03205522

Hock, H. S., & Schoner, G. (2010). Measuring perceptual hysteresis with the modified method of limits: Dynamics at the threshold. *Seeing and Perceiving, 23*(2), 173–195. doi:10.1163/187847510X503597

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604. doi:10.1073/pnas.0609485103

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *PNAS: Proceedings of the National Academy of Sciences, 106*(25), 10382–10385. doi:10.1073/pnas.0812142106

Jones, M., Mozer, M. C., & Kinoshita, S. (2009). Optimal response initiation: Why recent experience matters. *Advances in Neural Information Processing Systems, 21,* 785–792.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759–764. doi:10.1126/science.1169405

Kiger, J. I., & Glass, A. L. (1981). Context effects in sentence verification. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 688–700. doi:10.1037/0096-1523.7.3.688

Kleinschmidt, A., Buchel, C., Hutton, C., Friston, K. J., & Frackowiak, R. S. (2002). The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron, 34*(4), 659–666. doi:10.1016/S0896-6273(02)00694-3

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. doi:10.1037/0033-2909.119.2.254

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*(6), 2797–2822. doi:10.2307/1131753

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300. doi:10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Odic, D., & Halberda, J. (in press). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*.

Ludwig, C. J. H., & Davies, R. J. (2011). Estimating the growth of internal evidence guiding perceptual decisions. *Cognitive Psychology, 63*(2), 61–92. doi:10.1016/j.cogpsych.2011.05.002

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256–261. doi:10.1016/j.cognition.2011.07.009

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237. doi:10.1111/j.1467-8624.2011.01608.x

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*(3), 320–334. doi:10.1037/0097-7403.9.3.320

Michie, S. (1984). Why preschoolers are reluctant to count spontaneously. *British Journal of Developmental Psychology, 2*(4), 347–358. doi:10.1111/j.2044-835X.1984.tb00942.x

Newman, R. S., & Wick, P. L. (1987). Effect of age, skill, and performance feedback on children's judgments of confidence. *Journal of Educational Psychology, 79*(2), 115–119. doi:10.1037/0022-0663.79.2.115

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2012). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*. Advance online publication. doi:10.1037/a0029472

O'Leary, K. D., & O'Leary, S. G. (1977). *Classroom management: The successful use of behavior modification* (2nd ed., Vol. 17). Oxford, England: Pergamon.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., . . . Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41. doi:10.1016/j.cognition.2010.03.012

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547–555. doi:10.1016/j.neuron.2004.10.014

Pietroski, P., Lidz, J., Hunter, T., & Halberda, J. (2009). The meaning of "Most": Semantics, numerosity and psychology. *Mind & Language, 24*(5), 554–585. doi:10.1111/j.1468-0017.2009.01374.x

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864–901. doi:10.1037/a0019737

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107,* 1556–1563. doi:10.1152/jn.00985.2011

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333–367. doi:10.1037/0033-295X.111.2.333

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review, 7*(1), 1–25. doi:10.3758/BF03210723

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research, 49*(21), 2574–2585. doi:10.1016/j.visres.2009.06.009

Siegler, R. S., & Jenkins, E. A. (1989). *How children discover new strategies*. Hillsdale, NJ: Erlbaum.

Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 117–138. doi:10.1037/0278-7393.27.1.117

Tokita, M., & Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attention, Perception, & Psychophysics, 72*(7), 1839–1853. doi:10.3758/APP.72.7.1839

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science, 35,* 1518–1552. doi:10.1111/j.1551-6709.2011.01197.x

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*(3), 550–592. doi:10.1037/0033-295X.108.3.550

Vroom, V. H. (1964). *Work and motivation*. Oxford, England: Wiley.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*(5), 641–671. doi:10.1080/09541440802205067

Weinstein, Y., & Roediger, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition, 38*(3), 366–376. doi:10.3758/MC.38.3.366

Weinstein, Y., & Roediger, H. L. (2012). The effect of question order on evaluations of test performance: How does the bias evolve? *Memory & Cognition, 40,* 727–735. doi:10.3758/s13421-012-0187-3

Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision, 8*(5). doi:10.1167/8.5.7

Woodworth, R. S., & Schlosberg, H. (1972). *Experimental psychology*. Oxford, England: IBH Publishing.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1–B11. doi:10.1016/S0010-0277(99)00066-9

Appendix

Numerical Ratio Presented on Each Trial (e.g., 5 Blue vs. 15 Yellow, 15/5 = 3.00)

Trial	Easy-First	Hard-First	Trial	Easy-First	Hard-First
1	3.00	1.11	19	1.17	1.50
2	2.00	1.14	20	1.25	1.25
3	3.00	1.11	21	1.14	2.00
4	1.50	1.17	22	1.17	1.50
5	3.00	1.11	23	1.14	2.00
6	3.00	1.11	24	1.25	1.25
7	1.50	1.17	25	1.17	1.50
8	3.00	1.11	26	1.14	2.00
9	2.00	1.14	27	1.14	2.00
10	2.00	1.14	28	1.11	3.00
11	1.50	1.17	29	1.17	1.50
12	1.25	1.25	30	1.11	3.00
13	2.00	1.14	31	1.11	3.00
14	1.50	1.17	32	1.17	1.50
15	2.00	1.14	33	1.11	3.00
16	1.25	1.25	34	1.14	2.00
17	1.50	1.17	35	1.11	3.00
18	1.25	1.25			

Received April 24, 2012
Revision received October 2, 2012
Accepted October 11, 2012 ■