

On the Origins of Logarithmic Number-to-Position Mapping

Dror Dotan

INSERM-CEA Cognitive Neuroimaging Unit, Gif-sur-Yvette,
France, and Tel Aviv University

Stanislas Dehaene

INSERM-CEA Cognitive Neuroimaging Unit, Gif-sur-Yvette,
France, and Collège de France

The number-to-position task, in which children and adults are asked to place numbers on a spatial number line, has become a classic measure of number comprehension. We present a detailed experimental and theoretical dissection of the processing stages that underlie this task. We used a continuous finger-tracking technique, which provides detailed information about the time course of processing stages. When adults map the position of 2-digit numbers onto a line, their final mapping is essentially linear, but intermediate finger location show a transient logarithmic mapping. We identify the origins of this log effect: Small numbers are processed faster than large numbers, so the finger deviates toward the target position earlier for small numbers than for large numbers. When the trajectories are aligned on the finger deviation onset, the log effect disappears. The small-number advantage and the log effect are enhanced in dual-task setting and are further enhanced when the delay between the 2 tasks is shortened, suggesting that these effects originate from a central stage of quantification and decision making. We also report cases of logarithmic mapping—by children and by a brain-injured individual—which cannot be explained by faster responding to small numbers. We show that these findings are captured by an ideal-observer model of the number-to-position mapping task, comprising 3 distinct stages: a quantification stage, whose duration is influenced by both exact and approximate representations of numerical quantity; a Bayesian accumulation-of-evidence stage, leading to a decision about the target location; and a pointing stage.

Keywords: number representation, numerical estimation, mental number line, two-digit numbers

Supplemental materials: <http://dx.doi.org/10.1037/rev0000038.supp>

How do we understand the quantity represented by two-digit numbers? This question can be explored using a number-to-position mapping task, in which individuals are presented with a number and are asked to mark the corresponding position on a

number line. This task is informative because it requires converting the symbolic Arabic number into an internal representation of quantity. Quantities are thought to be internally organized along an internal continuum which has been likened to a mental “number line” (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Cappelletti, Kopelman, Morton, & Butterworth, 2005; Dehaene, Bossini, & Giraux, 1993; Fitousi & Algom, 2006; Ruiz Fernández, Rahona, Hervás, Vázquez, & Ulrich, 2011; Shaki, Fischer, & Petrusic, 2009; von Aster, 2000). Monitoring how participants point at the physical number line can therefore shed some light on this internal quantity representation (Barth & Paladino, 2011; Booth & Siegler, 2006; Dehaene, Izard, Spelke, & Pica, 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003).

To gain greater insight into this process, we developed a paradigm in which the participants performed the number-to-position task on a tablet computer while their finger position was continuously tracked (Dotan & Dehaene, 2013). The finger direction is assumed to continuously reflect the ongoing decision (Finkbeiner & Friedman, 2011; Finkbeiner, Song, Nakayama, & Caramazza, 2008; Freeman, Dale, & Farmer, 2011; Santens, Goossens, & Verguts, 2011; Song & Nakayama, 2008a, 2008b, 2009), and this allows for a temporal dissection of the digit-to-quantity conversion process. We observed that, starting at about 400 ms post stimulus onset, the finger position began to be correlated with the linear quantity of the two-digit target number. Most importantly, there was a transient time window, around 550–1,050 ms, in which the finger position was affected by an additional contribution of the

Dror Dotan, INSERM-CEA Cognitive Neuroimaging Unit, Gif-sur-Yvette, France, and Language and Brain Lab, School of Education, and the Sagol School of Neuroscience, Tel Aviv University; Stanislas Dehaene, INSERM-CEA Cognitive Neuroimaging Unit, and NeuroSpin Center and Department of Experimental Cognitive Psychology, Collège de France.

We thank Florent Meyniel for his help in the mathematical modeling; Ricardo Tarrasch and Pedro Pinheiro-Chagas for advising on the data analysis; and Rachel Somech, Dana Fuks, and Moria Lahis for their help in executing the experiments. This research was supported by INSERM, CEA, Collège de France, and a grant from the Bettencourt-Schueller Foundation. Dror Dotan is grateful to the Azrieli Foundation for the award of an Azrieli fellowship. Parts of the data reported here were presented in the Second Conference on Cognition Research of the Israeli Society for Cognitive Psychology, Akko, Israel (February 2016). Some of the data reported here (Experiment 7, Patient ZN) are new analyses on previously published data. Both authors conceived and designed the study and wrote the manuscript. Dror Dotan collected and analyzed the data. Both authors approved the final version of the manuscript for submission. The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Correspondence concerning this article should be addressed to Dror Dotan, Language and Brain Lab, School of Education, Tel Aviv University, Tel Aviv, Israel 69978. E-mail: dror.dotan@gmail.com

logarithm of the target. This observation suggested that the quantities were encoded by two distinct systems: an exact linear representation, where all numbers are equally well represented, and an approximate representation, where small numbers are represented more precisely than larger ones. This conclusion was in accord with studies that found compressive quantity representation in other tasks (Anobile, Cicchini, & Burr, 2012; Berteletti et al., 2010; Booth & Siegler, 2006; Dehaene et al., 2008; Dehaene & Marques, 2002; Lourenco & Longo, 2009; Núñez, Doan, & Nikoulina, 2011; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003; Viarouge, Hubbard, Dehaene, & Sackur, 2010). Mathematically, the approximate representation can be described as a logarithmic number line with fixed variance, as suggested by neural recordings and brain imaging data (Nieder & Miller, 2003; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). As previously noted (Dehaene, 1997), an equally accurate model of behavioral data can be obtained by postulating a linear number line with scalar variability (standard deviation proportional to number; for discussion, see Cicchini, Anobile, & Burr, 2014; Dehaene, 2007). As shorthand, we refer to these two representations simply as “approximate,” referring to the fact that they both show an increasing uncertainty as the numbers get larger.

Our goal in the present study was to clarify the theoretical reasons why a logarithmic effect arises even in adults, who know perfectly well that they should point to the linear location of the numbers. In particular, we designed new experiments exploring the hypothesis of a dual representation of quantity. We reasoned that, if there are two distinct representations of number, respectively exact and approximate, then we might be able to interfere with one of them and therefore transiently enhance the influence of the other. We relied on the method introduced by Anobile et al. (2012), who used quantity-to-position mapping in a dual-task setting. In the critical condition, participants estimated a number of dots and responded by marking a position on a line, while simultaneously performing a secondary task of color pattern judgment. This manipulation made their mapping more logarithmic. This pattern could be explained as a psychological refractory period (PRP) effect in which the secondary task competed with the exact linear quantification process for central resources, while leaving approximate quantification intact. As a result, the log effect was facilitated while the linear representation was reduced. The log-linear dissociation can therefore support a model of dual quantity representation. We aimed to replicate these findings with two-digit symbolic numbers, using our continuous number-to-position paradigm.

We also assessed a new theoretical interpretation that has recently arisen for the log effect in number-to-position tasks (Cicchini et al., 2014). This interpretation rests on a single quantity representation with differential variability—large quantities are represented with greater noise than small quantities. The idea is that the log effect results from a Bayesian process that combines this fuzzy quantity representation with prior knowledge (Fischer & Whitney, 2014; Jazayeri & Shadlen, 2010). Because large quantities are fuzzier than small quantities, they are estimated with lower confidence, and the Bayesian decision process assigns them a smaller weight relatively to prior knowledge. The decision is therefore slower (and the effect of prior biases is stronger) for larger target numbers than for smaller target numbers, and this is what gives rise to the logarithmic effect. In a dual-task setting, interference from the secondary task reduces even further the amount of evidence that can be extracted from the quantity

representation per unit of time, and therefore the logarithmic effect is increased.

Note that differential variability between small and large numbers can take many forms: One possibility is scalar variability (linear mapping of numerical quantities, and linear relation between the noise level and the target number), but the model can accept almost any form of differential encoding of small and large numbers. Thus, a compressive scale for number (e.g., logarithmic) with fixed variability would lead to similar results. Furthermore, when the stimuli are sets of dots (as was the case in Anobile et al., 2012), differential variability may arise from the assumption that the noise in the subitizing range (1–3) is lower than in the non-subitizing range (>4; Cicchini et al., 2014).

Crucially, according to this model, logarithmic mapping can be obtained even if the internal quantity scale is not logarithmic. Although it was initially argued that logarithmic behavior in the number-to-position task implies an internal logarithmic representation (Booth & Siegler, 2006; Dehaene et al., 2008; Dotan & Dehaene, 2013; Siegler & Booth, 2004; Siegler & Opfer, 2003), Cicchini et al.’s (2014) model shows that this is not the case. In particular, as previously argued, there is a near-complete behavioral equivalence between the log and the scalar variability models of approximate number representation (Dehaene, 2007).

Cicchini et al. (2014) further showed that the prior in the Bayesian decision process need not be fixed. Indeed, they discovered a new empirical finding that suggests that the prior is adjusted on a trial-by-trial basis: Judgments are strongly affected by the quantity presented on the immediately previous trial. Nevertheless, whether the prior is fixed or is updated after each trial, what really accounts for the log effect in Bayesian decision models is differential variability. Accordingly, a recent study has shown a logarithmic effect in quantity-to-position mapping even in the first trial of an experiment, when prior trial information was not yet available (Kim & Opfer, 2015).

The experiments and equations presented in Cicchini et al. (2014) capture only the participants’ ultimate response location in a number-to-position task, and remain silent about the sequence of processing stages that lead to this decision. In the present study, we wish to extend this model to account for the detailed within-trial dynamics of the number-to-position task. Our goal, indeed, is to obtain a detailed theory of the successive stages leading to a decision in the number-to-position task. We will show that an ideal-observer theory can account for our main finding that the mapping to position shows a logarithmic trend when the trial starts but becomes fully linear when the finger reaches the number line. The intuition behind this model can be specified succinctly: Assuming that the decision to move is based on a Bayesian decision process, with a progressive accumulation of evidence arising from the target, then differential variability should affect the *processing time* of the target. Large target numbers, which are represented with higher variability, are quantified more slowly than small target numbers (hereby, “small-number advantage”), so the Bayesian prior is overridden more slowly for large target numbers. As a result, at each poststimulus time point, small-target trials are in a more advanced stage of processing than large-target trials, which means that the finger trajectories for small targets are farther apart from each other than the trajectories for larger numbers. These differential distances between the trajectories appear as logarithmic effect when analyzing a specific time point.

We term this dynamic version of Cicchini et al.'s (2014) model *the differential encoding time model*. In the final section, we present a precise mathematical model and simulations of this idea. Note that the differential encoding time model conforms to the two main assumptions of Cicchini et al. (2014): (a) The target position is determined by a Bayesian decision process, with a prior that is affected by previous trials, and (b) the log effect results from differential variability for small versus large numbers, which causes differential overriding of the prior by the present-trial quantity.

Experiment 1: Number-to-Position Mapping With Dual Task

In Experiment 1, the participants mapped numbers between 0 and 40 to the corresponding positions on a number line. Each participant performed the task in three conditions, administered in three separate blocks. The first condition involved a single task: The participants mapped numbers to positions, with no other manipulation (like in Dotan & Dehaene, 2013).

The second condition involved dual-tasking: Subjects performed the number-to-position mapping parallel to a distracter task. Like Anobile et al. (2012), we used a color-detection task, which in our case was color naming. We hoped this task would maximize the interference effect, because it is not only attention demanding but also involves verbal output, which may selectively interfere with the linear quantity representation. One possible reason for such a selective interference rests on the assumption that Arabic numbers can be encoded as quantities by both hemispheres, but only the left hemisphere houses a verbal representation of numbers (Cohen & Dehaene, 2000). When the verbal system is occupied, approximate quantity may still be perceived without verbal mediation (Dehaene & Cohen, 1991; Dehaene et al., 2008; Dehaene, Spelke, Pineda, Stanescu, & Tsivkin, 1999). Under this hypothesis, verbal interference should increase the relative weight of the nonverbal parietal circuit that encodes approximate quantities. The dual-representation model thus predicts that the color naming condition should enhance the transient logarithmic effect.

The third (control) condition was number naming: The participants said aloud the number while pointing to the corresponding position. This condition does not divert attention from the target number and, if anything, should enhance the exact linear representation.

Method

Participants. Eighteen right-handed adults, aged $27;8 \pm 6;5$, with no reported learning disabilities or color blindness, were paid €10 for participation. Their mother tongue was Hebrew. For comparison, we also reanalyzed the data of 21 right-handed participants reported in Dotan and Dehaene (2013)—10 Hebrew speakers, nine French, one Italian, and one Thai, aged $35;5 \pm 10;7$, who performed the number-to-position task silently. Digital numbers in Hebrew are written like in English, and in our number-to-position paradigm Hebrew participants and left-to-right readers were found to exhibit similar patterns of results (Dotan & Dehaene, 2013).

Procedure. In each of the three conditions (silent, color naming, and number naming), each number 0–40 was presented six times (246 trials) in random order. In the two naming conditions, the participants were told that the two tasks (number-to-position and naming) were equally important but that they should first

attend to the naming task and then to the number-to-position task. The three conditions were blocked and were administered in random order (three participants per presentation order). The silent condition was identical to Dotan and Dehaene (2013): A horizontal number line, marked only with 0 and 40 in its ends, was present at the top of the screen throughout the experiment (see Figure 1). When the participants touched the initiation rectangle, a fixation cross appeared above the middle of the number line. When the participants started moving their finger toward the number line, the fixation cross was replaced by the target number and the participants moved their finger to what they judged to be the corresponding position on the number line. When the finger reached the number line, an arrow showed the position marked by the finger. In the naming conditions, while moving the finger the participants also said aloud (in Hebrew) the target number or a color name. One color per trial—white, yellow, orange, pink, red, blue, or green—was indicated by two horizontal stripes that appeared simultaneously with the target number, surrounding it. The oral responses were tape recorded and trials with semantic or phonological errors were excluded. The speech onset time was defined as the first time point in which the voice level, sampled at 20 Hz, exceeded a threshold level for a consecutive period of 200 ms. This threshold was configured per experiment session to match environment noise and the participants' speech volume.

The following violations were considered as failed trials: lifting the finger in midtrial, touching the screen with more than one finger, moving the finger backward, and starting a trial with sideways (rather than upward) movement. Furthermore, excluding the first 300 ms of each trial, a minimal finger velocity of 6 mm/s was required. The finger also had to reach the number line within 2 s and one third of the vertical distance within 1 s, with linear

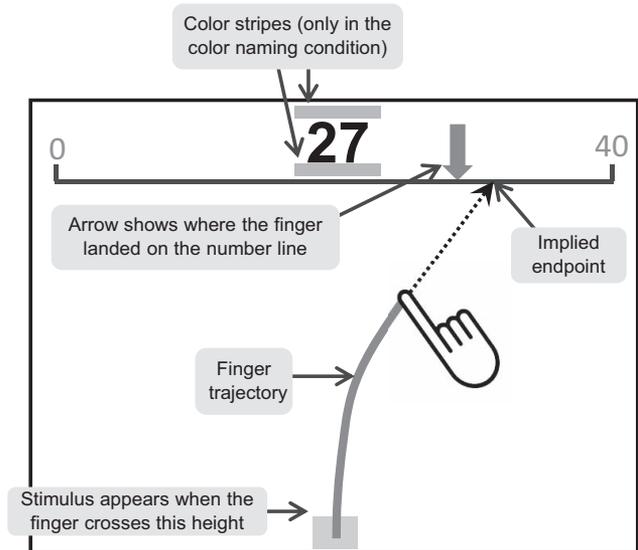


Figure 1. Task and screen layout. Participants pointed to the location of two-digit numbers on a horizontal number line that extended from 0 to 40. On each trial, they first placed their finger on the bottom rectangle. When they started moving their finger upward, the target appeared. In the naming conditions, they simultaneously named either the number itself (number naming block), or the color of the two stripes surrounding it (color naming block).

interpolation. In the two naming conditions, trials were also considered as failed if the speech onset time was too early (less than 200 ms) or too late (less than 200 ms before the finger reached the number line). Failed trials were excluded from all analyses and their target numbers were presented again later in the experiment. The training procedure was as described in Dotan and Dehaene (2013), with additional training phases for reviewing the color names and for adapting to the speech onset limits.

Technical specifications. We used Apple iPad 1 and iPad 2 devices in landscape orientation. The screen size was 197×148 mm and the resolution of display and finger tracking was $1,024 \times 768$ pixels. The background was black. The number line was white, 2×844 pixels, and horizontally centered 80 pixels below the screen top. Target numbers appeared centered above the number line in Arial bold white font with 10-mm-high digits. The numbers 0 and 40 at the ends of the number line appeared in light gray, 5-mm-high Helvetica font. The fixation cross was 7.7 mm high and wide. The feedback arrow was green, 7.7 mm high, and pointed downward with its tip touching the number line. In the color naming condition, colors were shown as two horizontal stripes (100×12 pixels each), horizontally centered, above and below the target number. The trial-initiation rectangle was 60×40 pixels (landscape) and dark gray. The stimulus appeared when the finger crossed the line $y = 50$ pixels (from the bottom of the screen).

Data encoding and preprocessing. We first introduce our terminology. The trial *endpoint* is the position where the finger crossed the number line (expressed in numerical units, 0–40). The *endpoint bias* is the difference between the endpoint and the target number. *Endpoint error* is the absolute value of endpoint bias. *Movement time* is the duration from stimulus onset to reaching the number line. *Horizontal movement onset time* is the post-stimulus-onset time in which we could identify that the finger started deviating toward the target number (see the “Identifying the onset of horizontal movement” section below).

The finger position was sampled at $60 \text{ Hz} \pm 1 \text{ ms}$, recorded as a sequence of time stamped x,y -coordinates, and transformed into a fixed sampling rate of 100 Hz using cubic spline interpolation. For each point along the trajectory, θ_t is the direction vector between the finger coordinates at times $t - 50$ ms and t . The θ_t values were smoothed using a Gaussian with $\sigma = 40$ ms. The *implied endpoint* (iEP) is the position on the number line that the finger would reach if it keeps moving in the direction θ_t . iEPs were cropped to the range $[-2, 42]$ and were undefined for sideways movement ($|\theta_t| > 80^\circ$).

A median trajectory was calculated per participant and target number as the median x,y -coordinates per time point (for late time points, which exceeded the trial’s movement time, the endpoint was used when averaging). iEPs of the median trajectories are denoted iEP_{med} .

We excluded from all analyses (and from the median trajectory calculation) the failed trials (as defined above), trials with movement time < 200 ms, trials with semantic or phonological naming errors, and trials with outlier endpoints. Outlier endpoints were defined, per group of six trials with the same target number, as the endpoints that exceeded the 25th or 75th percentile by more than 1.5 times the interquartile range.

Trajectory analysis. Trajectories were analyzed using the method introduced in Dotan and Dehaene (2013). One regression was run per participant and per time point in 50-ms intervals. The

dependent variable in these regressions was the iEP of the median trajectories (iEP_{med}). The predictors were the target number (denoted N_{0-40}), its logarithm, $\log'(N_{0-40}) = \log(1 + N_{0-40})$, linearly transformed to the range 0–40, the unit digit (U), and a bias function due to spatial reference points (SRPs, defined in Equation 1). The latter (or similar functions) was found significant in several number-to-position studies (Barth & Paladino, 2011; Dotan & Dehaene, 2013; Rouder & Geary, 2014; Slusser, Santiago, & Barth, 2013) and was hypothesized to result from the comparison of the target position with three spatial reference points—the middle of the number line (20) and its two ends (0 and 40).

$$\text{SRP}(N) = \begin{cases} 20 * \frac{\log(N+1)}{\log(N+1) + \log(21-N)} & \text{For } N \leq 20 \\ 20 + 20 * \frac{\log(N-20+1)}{\log(N-20+1) + \log(21-N)} & \text{For } N > 20 \end{cases} \quad (1)$$

A second-stage analysis was performed per predictor and time point, to examine whether the predictor had a significant group-level effect: The participants’ regression b values (significant and nonsignificant) were compared with zero using a t test. The reported p values are one-tailed when $\text{mean}[b] > 0$ and two-tailed when $\text{mean}[b] < 0$.

A possible concern about implied endpoints is whether they indeed reflect the subject’s intention in a given point in time. iEPs are calculated based on the momentary finger direction θ_t , which does not necessarily reflect a well-developed motor plan but perhaps random movements or jitter. However, in our previous publication with this paradigm we showed that analyzing iEPs yields similar results to an analysis of the x coordinates, which are much less affected by jitter—at least when it comes to regression analysis, in which all trials are pooled together. Here, we preferred the iEPs over the x coordinates, as they provide better temporal granularity (Dotan & Dehaene, 2013).

ANOVA. The speed of performing the number-to-position task varied a lot between individuals. Our goal in the present study was not to explain these interindividual differences, but to focus on the within-subject factors that affect people’s behavior in the number-to-position task. For this reason, in all analyses of variance (ANOVAs) in this study—most of which concern reaction time (RTs)—we use repeated measures design and report effect sizes as partial η^2 , a measure that is independent of the between-subjects variance. To maintain standardization, we also report η^2 for one-way ANOVAs, and generalized η^2 (Bakeman, 2005; Olejnik & Algina, 2003), denoted η^2_G , for ANOVA with several factors. In case of an ANOVA effect for which $df = 1$ and the effect direction has a clear prediction, we used the corresponding t test and one-tailed p values.

Results

General performance. Table 1 shows that number-to-position mapping was more difficult in the color naming condition than in the silent condition: The participants were less accurate (larger endpoint error), slower, and had more failed trials. Thus, the color naming manipulation was clearly effective. The number naming manipulation had a weaker effect: A smaller difference was observed in movement time and failed trial rate, and accuracy was similar to the silent condition. The participants’ unanimous subjective impression was that color naming was considerably harder than the two other conditions.

Table 1
General Performance Measures in Experiment 1

Measure	Silent	Color naming	Number naming
Failed trials (%)	3 ± 2.4	22.1 ± 8***	14.4 ± 10.3***
Invalid speech onset (%) ^a	—	10.6 ± 5.3	11.2 ± 9
Naming error (%) ^b	—	2.6 ± 2	.04 ± .13
Minimal velocity violation (%)	1.4 ± 2.1	7.6 ± 5.8***	1.8 ± 3.3
Other errors (%)	1.6 ± 1.6	1.3 ± .9	1.4 ± 1.3
Endpoint outliers (%)	4.6 ± 1.5	5.6 ± 1.9 [†]	4.7 ± 1.4
Movement time (ms)	1,102 ± 154	1,398 ± 131***	1,208 ± 132***
Endpoint bias (0–40 scale)	-.65 ± .45	-.68 ± .46	-.58 ± .39
Endpoint error (0–40 scale)	1.7 ± .42	2.1 ± .7***	1.74 ± .4
Speech onset time (ms)	—	898 ± 101***.c	695 ± 90

Note. The standard deviations refer to between-subject variance of the per-subject means.

^a Invalid speech onset: The verbal response was too slow, too fast, or no response was made. ^b Naming error: semantic or phonological. ^c Speech onset time was compared between the color naming and number naming conditions.

Paired *t* test vs. the silent condition: [†] one-tailed *p* < .1. *** *p* < .001.

The participants' median trajectories are presented in Figure 2a–c. The trajectory data was submitted to the two-stage regression analysis described above in Method. All four predictors showed significant effects in all conditions (Figure 3a–c and Table 2). The silent condition replicated the results from Dotan and Dehaene (2013), including the approximate time window for the significant effect of the log regressor (500–600 ms here, 450–750 ms in Dotan & Dehaene, 2013). The only essential difference was that here we did not observe an early contribution of the unit digit; instead, both decades and unit digits arose simultaneously as significant regressors, giving rise to a main effect of the linear value of the two-digit target number (see Figure 3). The pattern of significant effects in the two naming conditions was similar to the silent condition, but in the color naming condition the factors were observed in later time points, in accord with the slower finger movement in this condition.

Assessment of the dual representation model.

Color naming interferes with the linear factor and enhances the logarithmic factor. The regression *b* values of the log and linear factors in the silent condition were compared, per time point, versus the color naming condition using a paired *t* test (see Figure 3). This comparison confirmed the dissociation between the log and linear factors: As predicted, the color naming manipulation enhanced the log factor and reduced the linear factor. The linear factor in the

color naming condition was significantly smaller than in the silent condition from 450 ms to 850 ms, $b[N_{0-40}]_{\text{colors}} < b[N_{0-40}]_{\text{silent}}$, $t(17) > 1.75$, one-tailed *p* < .05. The pattern was reversed for the log factor: A significant difference, $b[\log'(N_{0-40})]_{\text{colors}} > b[\log'(N_{0-40})]_{\text{silent}}$, was observed from 650 to 750 ms, $t(17) > 2.1$, one-tailed *p* ≤ .05 (the difference $b[\log'(N_{0-40})]_{\text{colors}} < b[\log'(N_{0-40})]_{\text{silent}}$ from 500 ms to 550 ms did not reach significance, $t(17) < 1.4$, one-tailed *p* > .09). This dissociation supports the predicted enhancement of the approximate representation relative to the exact representation during dual-task interference.

This influence of color naming can be interpreted in two ways—either as facilitating the approximate quantity representation and weakening the linear representation, or as delaying the linear representation (i.e., the difference between the silent and color curves in Figure 3d can be viewed as either vertical or horizontal). If we accept the delay model, the delay size can be estimated from Figure 3d as ~50 ms around movement onset (~450–500 ms post stimulus onset), increasing to ~200 ms when crossing the *b* = 1 threshold (at ~670–870 ms). Figure 3e suggests that color naming may have slightly delayed the influence of the logarithmic factor too, but this delay was much smaller and never exceeded ~50 ms. The results are therefore compatible with the hypothesis that the linear quantity representation was delayed, which left the stage for the log representation to have a larger effect on the finger movement.

Table 2
Experiment 1: Time Windows (ms Post Stimulus Onset) in Which the Regression *b* Values Were Significantly Different From Zero (*p* ≤ .05)

Factor	Silent	Color naming	Number naming
$b[N_{0-40}] > 0$	450–end	200,500–end	450–end
$b[\log'(N_{0-40})] > 0$	500–600	550–700	500–600
$b[\log'(N_{0-40})] < 0$	750–end	1250–end ^a	850–end
$b[U] > 0$	750–end ^b	550, 1,150, 1,300–end	1,000–end
$b[\text{SRP}] > 0$	550–end	700–end	550–end
$b[\text{SRP}] < 0$	None	None	150–400

Note. U = unit digit; SRP = spatial reference points.

^a *p* < .05 in 1,250–1,450, 1,650–1,700, and .05 < *p* < .07 in the other time points. ^b *p* < .05 in 750–800, 900–1,000, 1,400–1,500, and .05 < *p* ≤ .08 in the other time points.

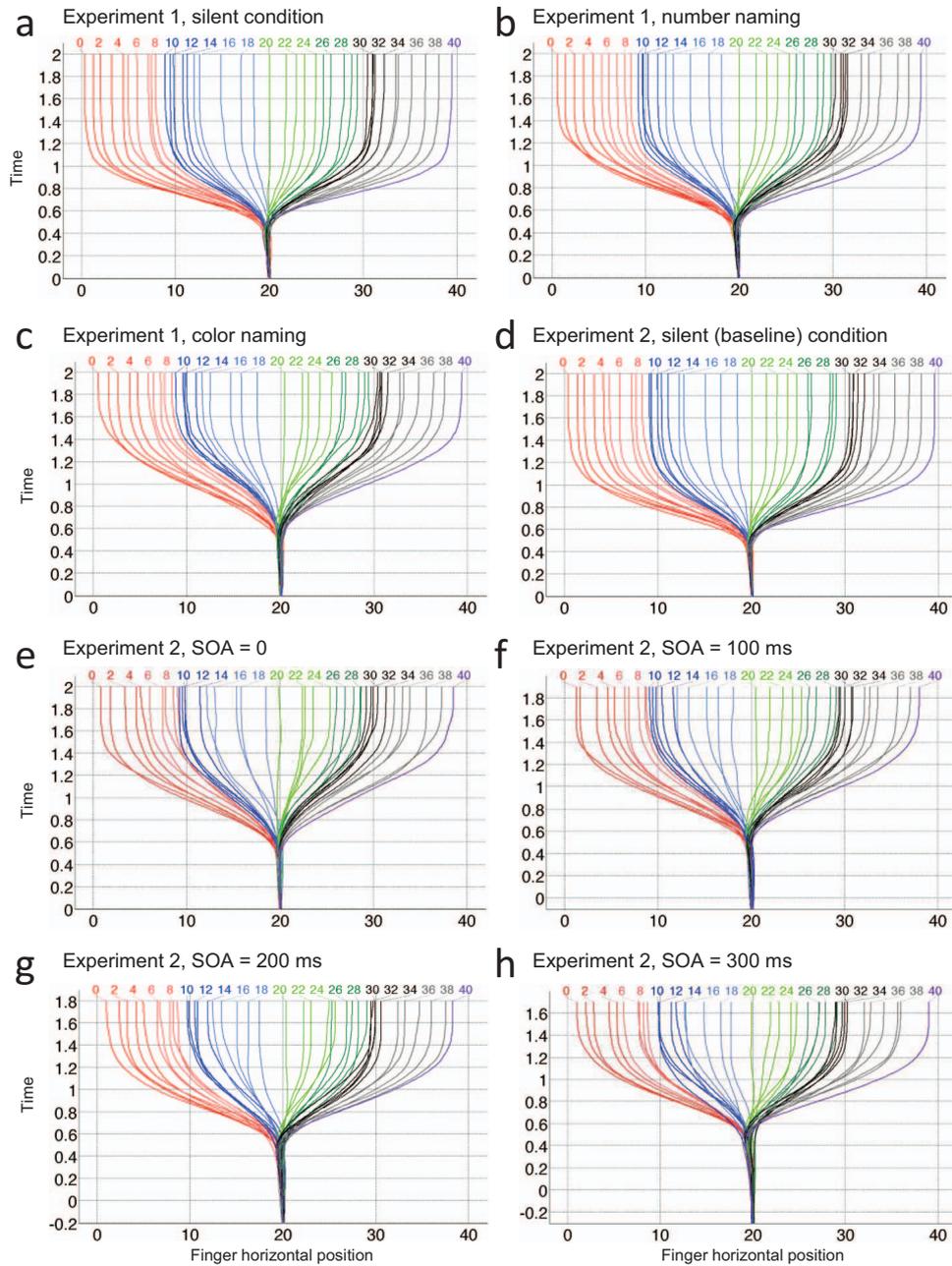


Figure 2. Median trajectories per target in Experiments 1 and 2. A median trajectory was created by resampling each trajectory into equally spaced time points, finding the per-subject median coordinates per time point, and averaging these medians over participants. Median trajectories shorter than 2 s were extended using the endpoint. Note that in Experiment 2, the number sometimes appeared after the color (Panels f–h); the bottom of each of these panels is aligned to the beginning of the trial (color onset), and time = 0 indicates the number onset. See the online article for the color version of this figure.

Given these apparent delays, we also attempted to compute a time-independent per-participant index of the peak log effect size. This index, denoted $b[\log'(N_{0-40})]_{\text{global}}$, was defined as the 75th percentile of $b[\log'(N_{0-40})]$ between 450 ms and 750 ms (the time window in which a significant log effect was found in Dotan & Dehaene, 2013; we used 75th percentile rather than

the peak b value to increase the robustness to noise and outliers). $b[\log'(N_{0-40})]_{\text{global}}$ was larger in color naming ($b = 0.20 \pm 0.10$) than in the silent condition ($b = 0.15 \pm 0.11$), $t(17) = 2.1$, one-tailed $p = .03$, Cohen's $d = 0.47$, confirming that the color naming manipulation enhanced the participants' log effect.

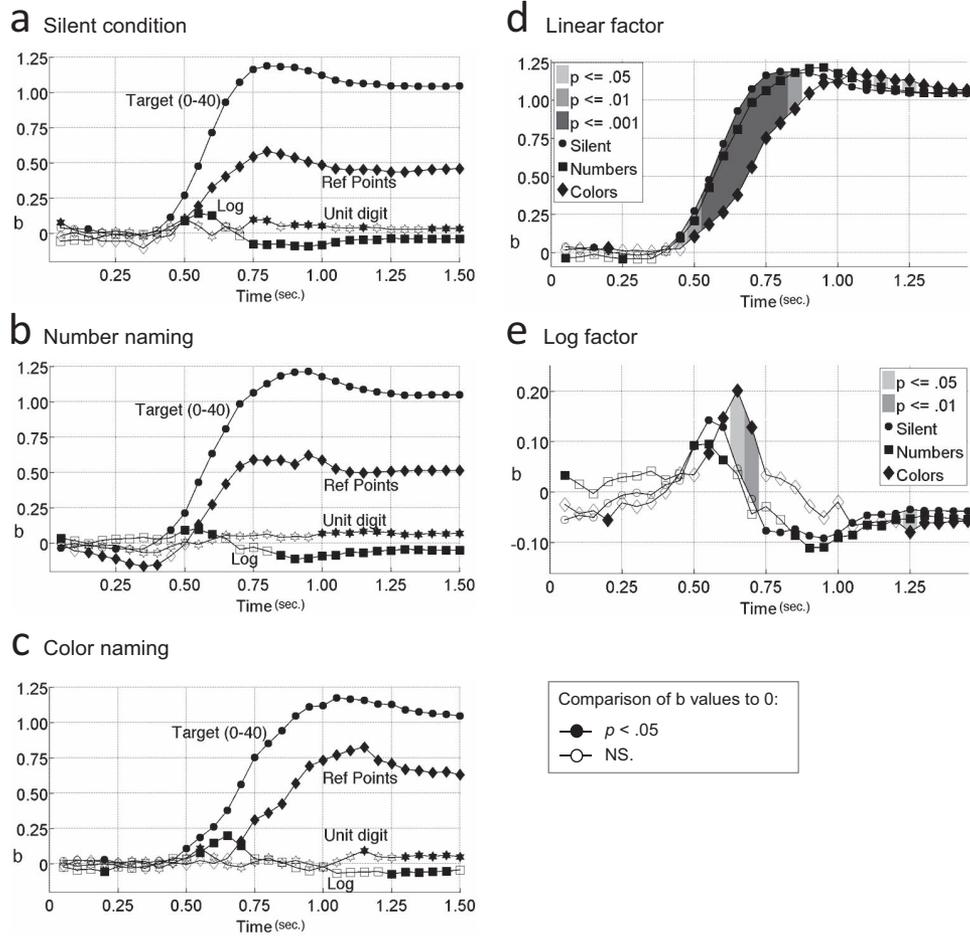


Figure 3. Time course of the effects in Experiment 1. All panels show b values of regressions on the implied endpoint of the median trajectory (iEP_{med}). One regression was run per time point, participant, and condition. The b values were averaged over participants and plotted as a function of time. In this and all subsequent regression figures, the b values were compared to zero (t test), and a black dot indicates a significant b value. (a-c) The b values per experimental condition. (d) The b values of the linear factor $b[N_{0-40}]$ in all three conditions: The effect of the linear factor arises faster in the silent condition than during color naming (the shaded area indicates a significant difference). (e) The b values of the logarithmic factor $b[\log'(N_{0-40})]$, showing a slightly stronger effect in color naming than in the silent condition.

When comparing regression b values in different conditions or at different time points, a potential confounding factor may be that the b values are affected by the global variance among trajectories $\sigma(iEP)$, and that this variance may differ between conditions. However, this explanation cannot account for the present results, because we found a *larger* log effect size in the color naming, whereas $\sigma(iEP)$ in equivalent time points was smaller in this condition. To completely rule out the alternative interpretation, we reran the log effect size analysis using the regression β values, which are not affected by the overall iEP s variance.¹ This analysis too showed a larger log effect size in color naming, $\beta[\log'(N_{0-40})]_{global/silent} = 0.22 \pm 0.22$, $\beta[\log'(N_{0-40})]_{global/colors} = 0.32 \pm 0.18$, $t(17) = 1.83$, one-tailed $p = .04$.

Control condition: Number naming. The number naming results were very similar to the silent condition (see Figure 3). The linear factor $b[N_{0-40}]$ was slightly smaller in number naming than

in the silent condition, but this difference was significant only at two time points (650 ms and 750 ms, $t(17) > 2.47$, two-tailed $p \leq$

¹ A regression analysis results in a regression formula $Predicted(y) = const + \sum \beta_i x_i$. These b values are informative when the predictors x_i and the dependent variable y are specified using a meaningful scale, as is the case in the present study. However, the b_i values are sensitive to the scale in which x_i and y are specified, and consequently they are typically not comparable with each other or across data sets. This comparability issue can be solved by standardizing the predictors and the dependent variable using linear transformation into a common scale with $M = 0$ and $\sigma = 1$. Denoting the transformed variables x'_i and y' , the regression formula would now be $Predicted(y') = \sum \beta'_i x'_i$, where $\beta'_i = b_i \times \sigma(x_i) / \sigma(y)$. Unlike b values, the β values are comparable with each other because all x'_i are specified using the same scale. More importantly for the present issue, even β values from different regressions are comparable, because the dependent variables too are specified using a fixed scale.

.03; in all other time points, $t(17) < 1.77$, $p > .09$). The log factor did not show a clear trend: $b[\log'(N_{0-40})]$ was stronger in the silent condition than in number naming at some time points and weaker at other time points, with a significant effect only in two time points (350 ms and 750 ms), $t(17) > 1.88$, $p < .04$. The global log effect size was similar in number naming, $b[\log'(N_{0-40})]_{\text{global}} = 0.12 \pm 0.13$, and in the silent condition, $b[\log'(N_{0-40})]_{\text{global}} = 0.15 \pm 0.11$; $t(17) = 1.12$, two-tailed $p = .25$. Thus, number naming, unlike color naming, did not facilitate the log factor. Analyzing the results in terms of delay shows that number naming caused only a small delay of ~ 10 – 20 ms in the linear factor and no delay in the log factor.

Dependency on prior trials. To assess the possibility that the participants' performance was affected by perseverations from previous trials, as described in Cicchini et al. (2014), the trajectory data was submitted to regression analysis with the four predictors described above, N_{0-40} , $\log'(N_{0-40})$, U, and SRP, to which we added the values of the target numbers in each of the last 15 trials (predictors denoted $N-1$, $N-2$, . . . , $N-15$). The regression was run on the raw, unaveraged trials and the dependent variable was iEP. One regression was run per condition and participant in 50-ms intervals. Per predictor, condition, and time point, the participants' b values were compared to zero using a t test (Figure 4a–c). These regressions showed a significant effect of the last two or three trials, which decreased around 500 ms as the finger began to point to the target quantity of the current trial.

To examine the relative effect of perseveration from each of the previous trials, we calculated the mean b value of each of the predictors $N-1$ to $N-10$ over the time range 0–600 ms. This was done for the three conditions in Experiment 1 and for the data from Dotan and Dehaene (2013). We observed an exponentially decreasing contribution of previous targets (Figure 4d). This pattern is consistent with the notion of a Bayesian process (Cicchini et al., 2014), according to which the finger is initially guided by an expectation or "prior" based on past trials, which gets constantly updated as the new target gradually overrides the expectation generated from older trials. The prior appears to decay roughly exponentially across trials $N-1$, $N-2$, $N-3$ and so forth, and in this respect, the phenomenon bears similarity to perseverations observed in many brain-lesioned patients (Cohen & Dehaene, 1998).

Assessment of the differential encoding time model. The differential encoding time model stipulates that the log pattern occurs because the finger deviates toward the desired location at an earlier time point for smaller target numbers (small-number advantage). As a result, in several post-stimulus-onset time points, trials with small target numbers are in a more advanced stage of processing (and finger movement) than trials with large targets, so the trajectories of small-target trials are farther apart from each other, giving rise to a log effect when regressing one post-stimulus-onset time point.

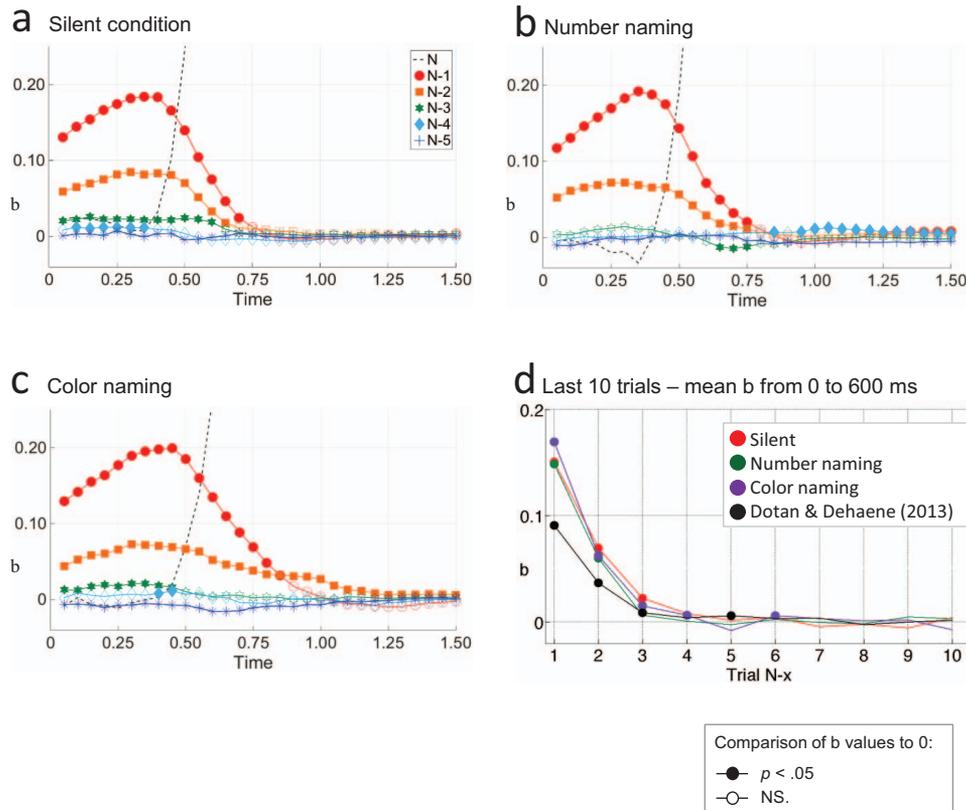


Figure 4. Influence of the prior targets on current finger trajectory in Experiment 1. (a–c) Influence of the current target N and the past five targets ($N-1$ to $N-5$) on the implied endpoint, as measured by regression (same type of plot as in Figure 3). (d) Mean b value over 0–600 ms, for each of the past 10 targets ($N-1$ to $N-10$), showing an exponentially decreasing influence of prior targets. See the online article for the color version of this figure.

Identifying the onset of horizontal movement. To assess the differential encoding time model, we first calculated the onset time of the finger's horizontal movement on each trial. To determine the horizontal movement onset per trial, we used an algorithm that aimed to identify the time point where the finger horizontal velocity started building up. A typical horizontal velocity profile of a trial consists of one or more velocity peaks (which may reflect several successive movement plans), but as every experimental measure it is also affected by jitter and random movements. Our goal was to find the onset of the earliest nonrandom velocity peak. To identify nonrandom peaks, we first estimated the participant's individual level of "motor noise" based on the distribution of horizontal velocities during the time window 0–250 ms (assuming that before 250 ms the movement is not yet affected by the target number; see Supplemental Materials for a justification of this assumption). We considered only velocity peaks that were significantly higher than this motor noise, and found the onset of the earliest of these peaks—as long as the onset occurred after 250 ms.

The specific algorithm was as follows. To calculate the horizontal velocity along each trajectory, we first applied Gaussian smoothing with $\sigma = 20$ ms to the finger x -coordinates, and then computed the derivative of the smoothed coordinates. To determine the horizontal movement onset per trial, we first looked for a significant peak of the x velocity profile—the highest x velocity that exceeded the top 1 percentile of the participant's velocity distribution on the first 250 ms of all trials. The onset time of this peak x velocity was defined as the latest time point where the x velocity remained lower than 5% of the peak velocity (if velocity never got below this threshold from 250 ms onward, no onset was found and the peak was ignored). To detect cases in which there was evidence for several successive movements (several velocity peaks), we checked if there was, earlier to the detected movement onset, another significant velocity peak, and reapplied the algorithm to detect this peak's onset. This procedure was applied recursively until no further velocity peak was detected. Visual inspection indicated that, for the vast majority of the trials, the algorithm was in excellent agreement with our subjective perception of the movement onset.

The algorithm failed to find the movement onset when the peak velocity was too low to reach significance, or when the 5%

criterion was never met in the time window from 250 ms post onset until 100 ms before the finger reached the number line. Such failures amounted to 19%, 18.3%, and 13.6% of the trials in the silent condition, number naming, and color naming, respectively. The horizontal movement onset time of these trials was coded manually whenever possible (the encoder was blind to the target number and saw only if it was smaller or larger than 20). After manual encoding, onset information was available for 97.9% of the trials. In the data of [Dotan and Dehaene \(2013\)](#), the algorithm failed to find the onset of 11.4% of the trials, and after manual encoding the onset information was available for 99.2% of the trials.

[Figure 5a](#) shows the mean horizontal movement onset times per target number and experimental condition. In the analyses of horizontal movement onsets (detailed below), we excluded trials with target number between 15 and 25, in which the target was close to the center of the screen and the horizontal movement was too small for reliable onset detection. We also excluded trials with endpoint outliers (as explained in the "Data encoding and preprocessing" section above).

The factors affecting the horizontal movement onset. The differential encoding time model predicts that the onset times should be earlier for smaller numbers (the small-number advantage effect), and that color naming should enhance the small-number advantage. To examine this assumption, the onset times were submitted to three-way repeated measures ANOVA with the subject as the random factor and with three within-subject factors: the experimental condition, the target side (<20 , left; or >20 , right), and a numeric factor given by the absolute distance between the target number and 20. Two separate ANOVAs were run: One compared color naming with the silent condition, and another compared number naming with the silent condition.

Color naming versus the silent condition. A significant main effect of condition, $F(1, 17) = 114.1$, $p < .001$, $\eta_p^2 = .87$, $\eta_G^2 = .38$, reflected the dual-task interference: Movement onset in color naming was delayed by 111 ms relative to the silent condition.

A significant main effect of side, $t(17) = 3.99$, one-tailed $p < .001$, $\eta_p^2 = .48$, $\eta_G^2 = .11$, confirmed the small-number advantage: Movement onset was earlier for small numbers than for large numbers (mean delay = 49 ms), as predicted by the differential encoding time model. The differential encoding time model also

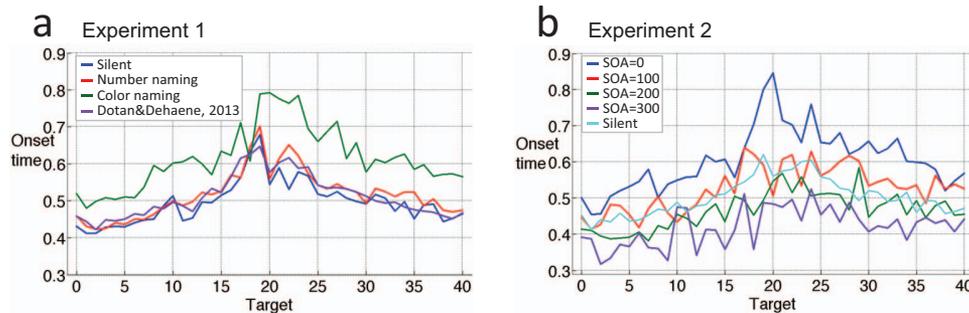


Figure 5. The mean onset time of horizontal movements, averaged over all participants, as a function of target number. (a) Onset time per condition in Experiment 1 and in [Dotan and Dehaene \(2013\)](#). (b) Onset time per SOA in Experiment 2 ($t = 0$ is the target number onset time). Targets = 15–25 are plotted here but were excluded from all analyses. See the online article for the color version of this figure.

predicts that color naming would facilitate the small-number advantage, and this was indeed the case: The small-number advantage in color naming (62 ms) was larger than in the silent condition (36 ms), and the Condition \times Side interaction was significant, $t(17) = 1.99$, one-tailed $p = .03$, $\eta_p^2 = .19$, $\eta_G^2 = .008$. Thus, the predictions of the differential encoding time model were fully confirmed.

A significant main effect of distance, $F(1, 17) = 97.4$, $p < .001$, $\eta_p^2 = .85$, $\eta_G^2 = .15$, showed that movement onset was earlier as the target number became closer to either end of the number line—a pattern clearly observable in Figure 5a. To analyze the interactions with the distance factor, we first examined their direction by calculating the distance effect in the various conditions. The movement onset time was submitted to regression analysis with distance = $|target - 20|$ as a single predictor—one regression per participant, condition, and side. The distance effect in color naming (average $b_{distance} = -9.92$ ms) was stronger than in the silent condition ($b_{distance} = -7.24$ ms). The three-way ANOVA showed that this difference was significant—Distance \times Condition interaction: $t(17) = 2.68$, one-tailed $p = .01$, $\eta_p^2 = .03$, $\eta_G^2 = .001$. The distance effect was also marginally stronger for numbers <20 ($b_{distance} = -9.49$ ms) than for numbers >20 ($b_{distance} = -7.67$ ms), Distance \times Side interaction: $t(17) = 1.43$, one-tailed $p = .08$, $\eta_p^2 = .11$, $\eta_G^2 = .001$. This Distance \times Side interaction is predictable by both logarithmic and scalar variability models, which attribute the distance effect to the target quantity: Such models predict a stronger distance effect when the ratios between the quantities are larger, as is the case for targets <20 compared with targets >20 . The three-way Condition \times Side \times Distance interaction was not significant, $F(1, 17) < 0.01$, $p = .98$.

Number naming versus the silent condition. A significant main effect of condition, $F(1, 17) = 12.43$, $p = .003$, $\eta_p^2 = .42$, $\eta_G^2 = .02$, reflected a dual-task interference, although smaller than in color naming: Movement onset in number naming was delayed by 18 ms relative to the silent condition.

A significant small-number advantage was observed (41 ms), main effect of side: $t(17) = 3.14$, one-tailed $p = .005$, $\eta_p^2 = .37$, $\eta_G^2 = .10$. The small-number advantage did not differ significantly between number naming (45 ms) and the silent condition (36 ms), Condition \times Side interaction: $F(1, 17) = 0.53$, $p = .48$.

The main effect of distance was significant, $F(1, 17) = 81.7$, $p < .001$, $\eta_p^2 = .83$, $\eta_G^2 = .13$ and this effect too did not interact with condition, $F(1, 17) = 0.45$, $p = .51$. The direction of the Distance \times Side interaction was examined using the same method we described above to analyze the color naming condition. This analysis showed that as predicted, the distance effect for numbers <20 ($b_{distance} = -9.0$ ms) was marginally larger than the distance effect for numbers >20 ($b_{distance} = -7.24$ ms), Distance \times Side interaction: $t(17) = 1.64$, one-tailed $p = .06$, $\eta_p^2 = .14$, $\eta_G^2 = .004$. The three-way Condition \times Side \times Distance interaction was not significant, $F(1, 17) = 0.28$, $p = .60$.

Differential encoding times as the reason for the log effect. The differential encoding time model attributes the transient log effect (see Figure 3) to earlier horizontal movement onset times in small-target trajectories than in large-target trajectories. If these differences in movement onset times were eliminated, the model predicts that the transient log effect would disappear. To eliminate onset time differences, we aligned each trial's trajectory data to its horizontal movement onset time. The aligned trajectories (exclud-

ing trials with no movement onset information) were submitted to regression analysis similar to the one described in the "Assessment of the dual representation model" section above, with iEP as the dependent variable and with four predictors: N_{0-40} , $\log'[N_{0-40}]$, the unit digit U, and SRP. One regression was run per condition, participant, and post-horizontal-movement-onset time point in 50-ms intervals. Per predictor, condition, and time point, the participants' b values were compared with zero using a t test. A significant positive contribution of $b[N_{0-40}]$ was found in all conditions and in all time points (see Figure 6). $b[N_{0-40}]$ was significant even at the time of horizontal movement onset ($t = 0$), and within 50 ms it reached a considerable effect in all conditions (over participants, mean $b > 0.38$). This indicates that when the finger horizontal movement started, the participants already had a linear quantity representation of the two-digit number. Crucially, the log factor $b[\log'(N_{0-40})]$ no longer showed any significant positive effect in any experimental condition, excluding a short time window (150–250 ms) in the Dotan and Dehaene (2013) data, in which there was a minor log effect, $b[\log'(N_{0-40})] \leq 0.05$ (Figure 6d). Thus, controlling for the movement onset time eliminated the log effect, as predicted by the differential encoding time model.

The elimination of the log effect cannot be attributed to the fact that the aligned regression was run only on a subset of the trials (those for which we could identify the movement onset): When the same regression was run on the same subset of trials without aligning trajectories by their onset time, the log factor $b[\log'(N_{0-40})]$ was significantly larger than zero in each of the 3 conditions during at least 250 ms, with peak $b[\log'(N_{0-40})] \geq 0.11$ (average over participants). These findings indicate that the transient log effect in this task, both in the silent single-task condition and in the dual-task conditions, can be fully explained by differential horizontal movement onsets per target.

Discussion of Experiment 1

The silent condition in Experiment 1 replicated the results of Dotan and Dehaene (2013): The analysis of trajectories showed a strong linear effect and a transient logarithmic effect. The color naming condition confirmed the prediction that dual-tasking makes the number-to-position mapping more logarithmic: The regression analysis showed a decreased (or delayed) linear factor and an enhanced log factor. This is similar to the results previously found when the quantities were presented nonsymbolically (Anobile et al., 2012).

The log-linear dissociation was initially taken as direct evidence for separate log and linear quantity representations (Dotan & Dehaene, 2013), with the linear representation being more sensitive to interference from the dual task—presumably due to competition of resources between the color naming task and linear quantity encoding mechanisms. However, the analysis of movement onsets suggests a simpler explanation: The decision to start moving the finger horizontally is earlier for smaller target numbers, thus the trajectories fan out more quickly for smaller number than for larger numbers, and this induces a transient log effect in the regressions. The dual task (color naming) further enhances this differential delay in movement onset as a function of target size, and consequently increases the log effect. Thus, the differential movement time model fully accounts not only for the dissociation

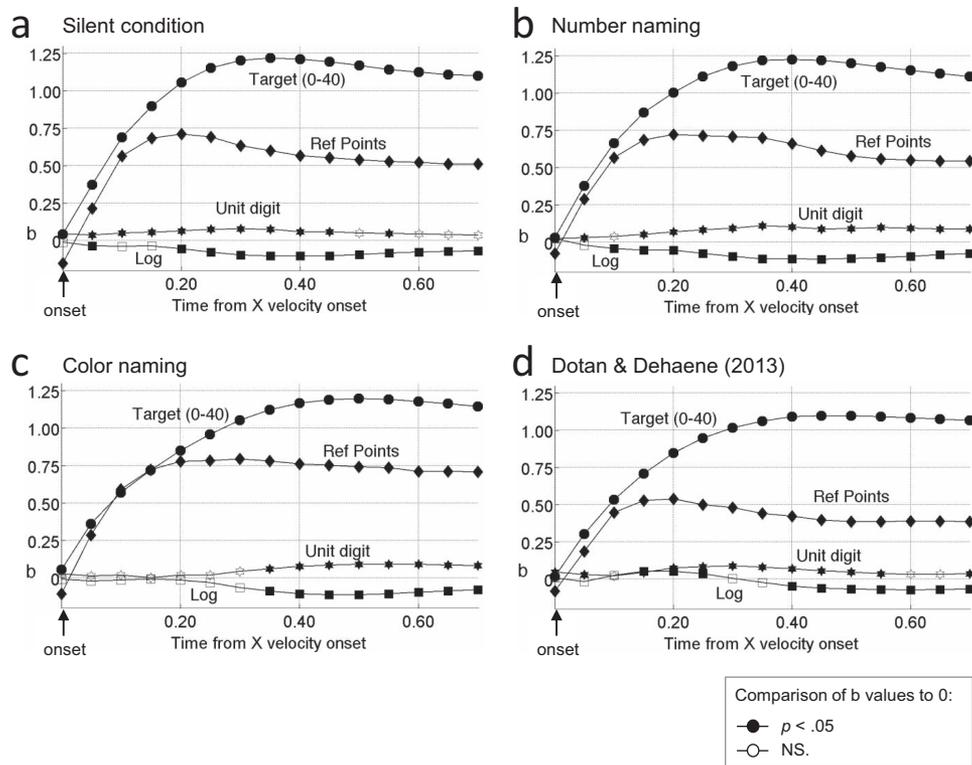


Figure 6. Time course of the effects in Experiment 1 after alignment on horizontal movement onset time. The figure shows the regression b values (dependent variable: implied endpoint) per condition and time point in Experiment 1 and, for comparison purposes, in *Dotan and Dehaene (2013)*, averaged over participants. The x -axis indicates the time after the initiation of horizontal movement.

between the silent and color naming conditions, but also for the log effect in each of the conditions. Indeed, when the horizontal movement onset time was controlled for (by aligning each trial to its movement onset time), the log effect was eliminated, and with it the difference between the conditions.

Our findings are consistent with the idea that, early in the trial, before the participants obtain evidence from the target, they move their finger in accordance to prior knowledge. In our task, participants initially point toward the midpoint of the line, which happens to be the optimal prior given the flat distribution of target numbers. Furthermore, their pointing is also influenced in part by the distribution of previous targets: When the previous targets are large, the finger is slightly displaced toward the right side, and vice versa. This effect is essentially a replication of *Cicchini et al.'s (2014)* finding of a prior-trial effect, although in our case the effect (a) showed an exponentially decreasing influence of several recent targets, and (b) influenced only the initial part of the next trial's finger trajectory, not the final endpoint.

The aligned-by-movement-onset analysis also showed that the unit and decade digits affected finger movement in an almost accurate 1:10 ratio throughout the trial, indicating that the decade and unit quantities were assigned accurate relative weights. This finding is interesting because it suggests parallel rather than sequential processing of the two digits: If one of the digits was processed before the other, its effect on movement should have been larger than implied by the 1:10 ratio. The absence of such

deviation from the 1:10 ratio suggests either that the decade and unit quantities were processed in parallel, or that the decision to initiate finger movement was delayed until a complete two-digit quantity was constructed.

Last, the analysis of movement onsets revealed a strong distance effect that was not predicted by any of the models: The movement onset was much earlier for targets close to the ends of the number line and delayed for targets near the middle (see *Figure 5*). The origins of this effect are discussed in Experiment 5.

Methodologically, these results indicate that data from the number-to-position paradigm should be interpreted with caution. Regression analyses of stimulus-aligned finger trajectories, as performed in our earlier publication (*Dotan & Dehaene, 2013*), show log and linear patterns at different times, yet this does not necessarily reflect directly the underlying internal representations. Rather, movement-aligned analysis suggests that this pattern may reflect the differential durations of a premovement stage of intention buildup.

An alternative interpretation of the small-number advantage is in terms of a motor rather than a numeric effect. According to this interpretation, the faster deviation to small numbers would not result from their magnitude but from their location on the left side on the number line. Purely motoric reasons, including for instance the types of muscle activity required to push the finger left or right, may make leftward movements faster than rightward movements. We refuted this hypothesis, however, with two control experi-

ments, which are reported fully in the supplemental material. In one experiment, the silent condition of Experiment 1 was replicated with a group of left-handed participants. The task required these participants for the cognitive operation as in Experiment 1, but for a reversed muscle operation. The motor hypothesis therefore predicts that the left-handed participants would deviate more quickly toward the right side (large numbers), that is, a large-number advantage. However, the findings were exactly the opposite: The left-handed participants showed a small-number advantage just like the right-handed group. In a second control experiment, a group of right-handed participants pointed to the same 41 locations as in the number-to-position task, but the target location was now indicated explicitly and non-numerically by an arrow placed at the target location. Thus, the set of required responses in this task was as in Experiment 1, but the decision process did not involve numbers. These participants deviated slightly faster to the right than to the left, that is, the opposite of the bias we observed in the numerical experiments. Taken together, these control experiments clearly refute the motor hypothesis and support our interpretation of the small-number advantage as a numerical effect.

Experiment 2: Manipulating the Color-Number SOA

Experiment 2 was designed to replicate the dual-task interference effect observed in Experiment 1 within the better-controlled setting of a PRP design (Pashler, 1984, 1994). In Experiment 1, the three conditions were very different from each other: One condition was a single task, and the two other conditions were dual-tasks involving naming of words from different categories (numbers and colors), which could trigger different cognitive processes (Bachoud-Levi & Dupoux, 2003; Bormann, Seyboth, Umarova, & Weillera, 2015; Cohen, Verstichel, & Dehaene, 1997; Dotan & Friedmann, 2015; Marangolo, Nasti, & Zorzi, 2004; Marangolo, Piras, & Fias, 2005). Experiment 2 therefore used the classic PRP manipulation of stimulus-onset asynchrony (SOA) between two fixed tasks. Only the color naming task was used, but the SOA between the onset of the color and the target number was manipulated. We assumed that decreasing the SOA would increase the temporal overlap between the central decision stages of the two tasks, thus imposing a decision bottleneck (Sigman & Dehaene, 2005). Thus, the effect of shortening the SOA would be similar to adding the dual-task in the first place. Consequently, we predicted an increased log effect for shorter SOAs, which according to the differential encoding time model should be entirely reducible to a

differential delay of movement onset for different numerical targets.

Method

Twenty right-handed adults (age $26;2 \pm 4;0$) with no reported cognitive deficits or color blindness were paid €10 for participation. Their mother tongue was Hebrew.

One experimental block was a replication of the color naming condition in Experiment 1. In three other blocks, the color stripes still appeared when the finger started moving, but the onset of the target number was delayed by 100 ms, 200 ms, or 300 ms. Each participant performed all blocks and was randomly assigned to one of four block presentation orders (0–100–200–300, 100–0–300–200, 300–200–100–0, or 200–300–0–100). Each number between 0 and 40 was presented twice per block (82 trials). The participants also performed silent number-to-position mapping (identical with the silent condition in Experiment 1) as a fifth block, which was administered last and presented each number four times (for five participants) or six times (for the other participants).

The horizontal movement onset time was calculated per trial using the method described above (“Identifying the onset of horizontal movement” section), excluding trials with target numbers 15–25. The automatic algorithm succeeded finding the onset of 90.8% of the trials (88.9%, 89.9%, 91.6%, and 92.6% per SOA condition) and 84.5% of the trials in the silent control condition. For the remaining trials, horizontal movement onset was encoded manually, after which 98.7% of the trials (and 98.6% of the control trials) had movement onset information. The other trials were excluded from the onset-related analyses. The onset analyses described below were also run while excluding the trials with manual onset encoding, and the results were essentially the same.

Results

Comparison of the conditions using trial-level measures. Table 3 shows the basic performance measures in this experiment. Each of these measures was compared across the four SOA conditions using repeated measures ANOVA with the per-subject mean as the dependent variable. There were no significant differences between the SOAs in endpoint bias, $F(3, 57) = 1.67, p = .18$ and endpoint error, $F(3, 57) = 1.66, p = .19$, but there were differences in movement time, $F(3, 57) = 3.65, p = .02, \eta_p^2 = .16, \eta^2 = .03$, and failed trial rate, $F(3, 57) = 3.98, p = .01, \eta_p^2 = .17$,

Table 3
General Performance Measures in Experiment 2

Measure	Silent	0 ms	100 ms	200 ms	300 ms
Failed trials (%)	2.6 ± 2.0	19.6 ± 10.2	21.4 ± 10.7	16.5 ± 10.8	14.4 ± 12.3
Movement time (ms) ^a	1,180 ± 162	1,414 ± 152	1,410 ± 143	1,428 ± 129	1,476 ± 155
Endpoint bias (0–40 scale)	–.66 ± .46	–.67 ± .72	–.57 ± .56	–.61 ± .67	–.78 ± .92
Endpoint error (0–40 scale)	1.7 ± .43	2.49 ± 1.1	2.41 ± .82	2.27 ± .98	2.43 ± 1.16
Speech onset time (ms) ^a	—	878 ± 106	871 ± 140	811 ± 138	802 ± 153
Horizontal movement onset time (ms) ^b	496 ± 42	597 ± 86	525 ± 83	467 ± 75	420 ± 75

Note. Standard deviations refer to between-subject variance of the per-subject means.

^a The movement time and the speech onset time are indicated with respect to the color onset time. ^b The horizontal movement onset time is indicated with respect to the number onset time.

$\eta^2 = .06$. The results were essentially the same when the ANOVA was run with the SOA as a numeric factor.

We continued with a classical PRP analysis, which consists in examining how the RTs in the two tasks were affected by the SOA manipulation. The two RT measures are the speech onset time for the naming task and the movement onset time for the pointing task.

The speech onset times of color naming were significantly different between the SOA conditions, one-way repeated measures ANOVA, $F(3, 57) = 10.9, p < .001, \eta_p^2 = .36, \eta^2 = .06$. They were longer in SOA = 100 than in SOA = 200, paired $t(19) = 4.0$, one-tailed $p < .001$, Cohen's $d = 0.10$, but were similar between SOAs 0–100 and 200–300, paired $t(19) < 0.54$, one-tailed $p > .6$.

The horizontal movement onset time too was significantly different between the SOA conditions, one-way repeated measures ANOVA, $F(3, 57) = 91.93, p < .001, \eta_p^2 = .83, \eta^2 = .42$; for all pairs of adjacent SOAs, paired $t(19) \geq 3.0, p < .001$, Cohen's $d > 1.1$. Table 3 shows that each increase of the SOA by 100 ms decreased the horizontal movement onset time by ~50–70 ms. Note, however, that this added delay is significantly smaller than the 100-ms spacing of the SOA conditions: when comparing onset times after adding 100 ms to the earlier SOA: paired $t(19) > 2.56, p < .02$, Cohen's $d > 0.57$ for all adjacent SOAs. In many PRP experiments, a 1:1 relation between SOA shortening and secondary-task delay is obtained (Pashler, 1984, 1994; Sigman & Dehaene, 2005). The fact that it was not obtained here suggests that interference was not complete and that there was partial resource sharing (Tombu & Jolicoeur, 2002) or intertrial variability in the prioritizing of the two tasks, as also confirmed by the above finding that color naming too was significantly delayed by shortening the SOAs.

Regression analysis of the trajectories. The trajectory data was submitted to regression analysis with iEP as the dependent variable and with the predictors introduced in Experiment 1: N_{0-40} , $\log'(N_{0-40})$, the unit digit U, the spatial-reference-points-based bias function SRP, and the target number of the previous trial, $N-1$. One regression was run per SOA (and for the silent condition), participant, and time point, in 50-ms intervals. The per-subject regression b values of each SOA, time point, and predictor were compared versus zero using a t test. The pattern of factors we observed in Experiment 1 was replicated for all four SOAs (Figure 7a–d): dominant linear factor, transient logarithmic factor, SRP contribution in the late trajectory parts, and an effect of the previous trial in early trajectory parts.

We then examined the effect of SOA on the linear factor (Figure 7e). The per-subject $b[N_{0-40}]$ values were first compared using a repeated measures ANOVA with a factor of SOA (one ANOVA per time point, starting from 150 ms). A significant difference between SOAs was found from 550 ms to 900 ms, $F(3, 57) > 4.48, p < .01, .19 < \eta_p^2 < .34, .03 < \eta^2 < .11$. A comparison of $b[N_{0-40}]$ between each pair of adjacent SOAs using paired t test showed that for SOAs from 0 to 200 ms, the difference was in the predicted direction, that is, decreasing the SOA resulted in a reduced linear factor. We found a significant difference $b[N_{0-40}]_{SOA = 100} > b[N_{0-40}]_{SOA = 0}$ from 700 ms to 900 ms, $t(19) > 1.89$, one-tailed $p < .04, 0.42 < \text{Cohen's } d < 0.56$, and $b[N_{0-40}]_{SOA = 200} > b[N_{0-40}]_{SOA = 100}$ from 550 ms to 750 ms, $t(19) > 1.82$, one-tailed $p < .05, 0.41 < \text{Cohen's } d < 0.74$. There was no significant difference between SOAs 200 ms and 300 ms at any time point, $t(19) < 0.84$, one-tailed $p \geq .21$; in fact, as Figure 7e clearly shows, the linear

factor was almost identical for these two SOA values. Thus, as predicted, decreasing the SOA (and thereby extending the time overlap between the two tasks) caused an increasing interference with the linear factor of the number-to-position task, which can be interpreted as a delayed onset of this factor. The shape of this effect, with an absence of a difference between the longer SOAs (200 and 300 ms), is classical for the PRP effect (Pashler, 1984, 1994; Sigman & Dehaene, 2005). It suggests that the central competition between the two tasks lasted no more than 200 ms, and therefore reached a floor level for SOA of 200 ms and beyond.

The effect of SOA on the log factor was examined in a similar manner (Figure 7f). No significant SOA effect on $b[\log'(N_{0-40})]$ was found in any time point: a per-time-point repeated-measures ANOVA, starting from 150 ms, with SOA as a within-subject factor and the subject as the random factor, showed no significant difference, $F(3, 57) < 2.15, p > .10$. Thus, whereas in Experiment 1 we observed significant effects on both the log and linear factors but in opposite directions, in Experiment 2 shortening the SOA reduced the linear factor while keeping the log factor almost unchanged.

The interaction between the log and linear factor was evaluated using two-way repeated measures ANOVA with the regression b values as the dependent variable, between-subjects factors of regression predictor (log, linear) and SOA, and the subject as the random factor. One ANOVA was run per time point, starting from 150 ms. A significant interaction was found from 600 ms to 850 ms, $F(3, 57) > 3.16, p \leq .03$, except $p = .06$ in time point 650 ms; $.12 < \eta_p^2 < .16, .02 < \eta^2 < .06$, confirming that the SOA manipulation affected the linear and log factor differently.

We assumed that the effect of prior from the previous trial would initially be independent of the new number presented, and consequently independent of SOA. Indeed, when aligning the SOA conditions to the beginning of the trial, that is, to the color onset rather than to the number onset (Figure 7h), no significant differences in $b[N-1]$ were found between SOAs until 550 ms (repeated measures ANOVA per time point, with $b[N-1]$ as the dependent variable, SOA as a within-subject factor, and the subject as the random factor), $F(3, 57) < 1.63, p > .19$. In later time points, from 600 ms to 900 ms (the downhill part of the $b[N-1]$ curve), a significant difference was found between the SOA conditions, from 600 ms to 900 ms, $F(3, 57) > 3.15, p \leq .03, .14 < \eta_p^2 < .42, .06 < \eta^2 < .18$; between 650 ms and 850 ms, $F(3, 57) > 5.51, p \leq .002$. This late between-SOA difference almost disappeared when the conditions were aligned to the number onset rather than to the color onset (Figure 7g), from 400 ms to 1,000 ms, $F(3, 57) > 2.44, p > .07$, except two time points, 650–700 ms, in which $p = .05$. Thus, the initial effect of $b[N-1]$ was triggered by the color onset, whereas its decay was linked to the number onset. These findings suggest that finger movement is initially affected by the prior from previous trial(s), and this effect decays as the prior is overridden by the new number presented.

Differential encoding times as the reason for the log effect.

The differential encoding time model assumes that the log effect occurs because the horizontal movement onset time is different for different target numbers. Once these onset differences are eliminated by aligning trajectories to their movement onset, the regression analysis should show no logarithmic effect. To examine this prediction, the trajectory data was submitted to regression analysis after aligning each trajectory to the trial's horizontal movement

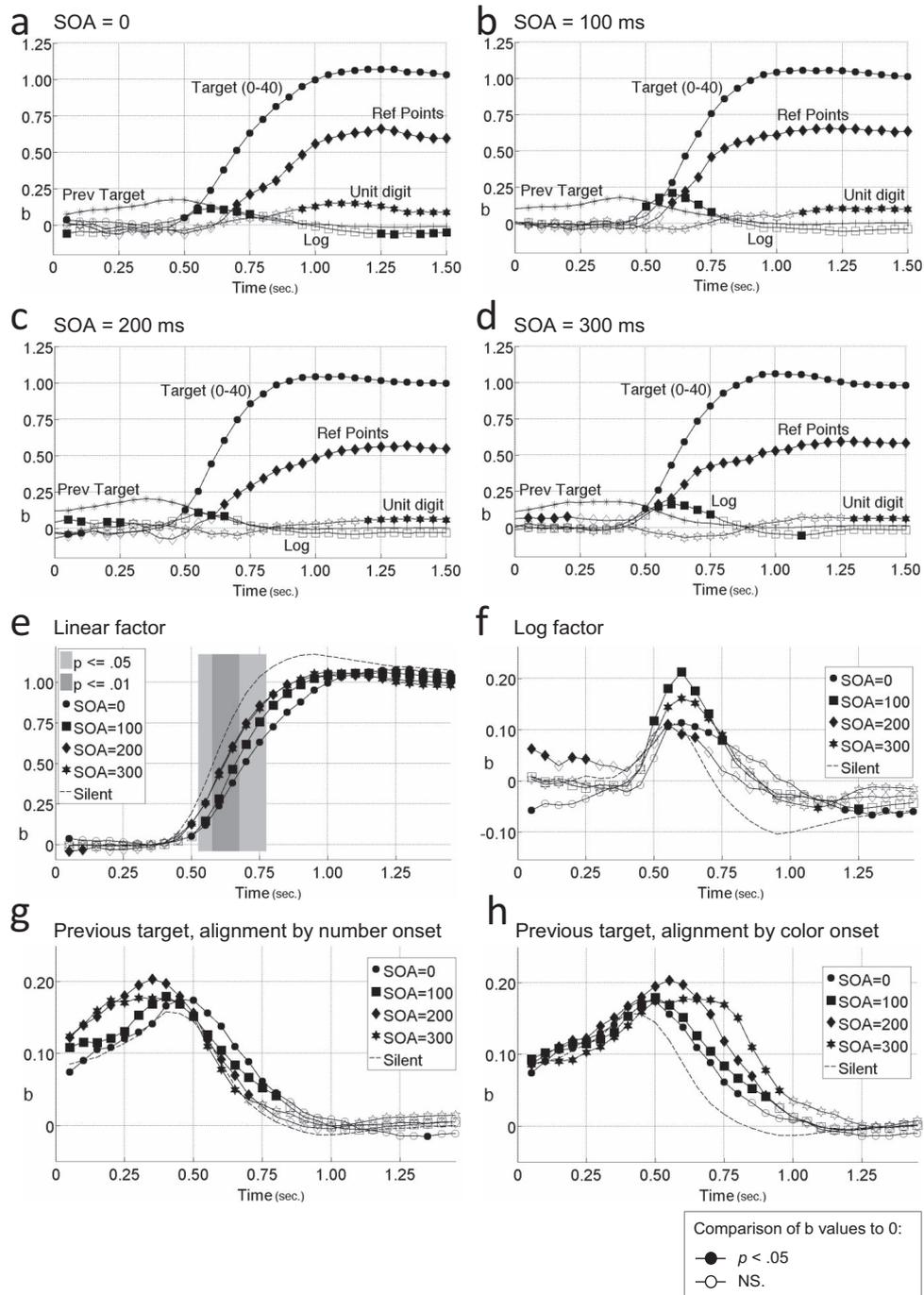


Figure 7. Time course of the effects in Experiment 2. Note that the different experimental conditions are horizontally aligned to the target number onset, not the color onset. Each of the panels (a–d) shows the regression factors of one SOA. (e) The linear factor $b[N_{0-40}]$ per SOA. Gray areas show a time window of 200 ms during which $b[N_{0-40}]_{SOA=100} < b[N_{0-40}]_{SOA=200}$. A similar difference $b[N_{0-40}]_{SOA=0} < b[N_{0-40}]_{SOA=100}$ was found in a slightly later time point, 700 ms to 900 ms. (f) The log factor $b[\log'(N_{0-40})]$ showed no significant differences among SOAs. (g–h) The prior-target factor $b[N-1]$, with the SOA conditions aligned by the (g) number onset or by the (h) color onset. The prior effect is initially independent of the number onset time, but its decay is linked to the number onset.

onset time. The dependent variable was iEP and the predictors were N_{0-40} , $\log'(N_{0-40})$, the unit digit U, and SRP. One regression was run per SOA, participant, and time point in 50-ms intervals. Per predictor, SOA, and time point, the participants' b values were compared with zero using a t test (see Figure 8). The linear factor $b[N_{0-40}]$ in these regressions showed a virtually identical pattern for all SOAs (Figure 8a). A per-time point repeated measures ANOVA, with SOA as a single within-subject factor and the subject as a random factor, showed no difference in $b[N_{0-40}]$ between SOAs at any time point from 50 ms, $F(3, 57) < 2.21$, $p > .09$, and only a minor difference at $t = 0$, $F(3, 57) = 3.7$, $p = .02$, $\eta_p^2 = .16$, $\eta_c^2 = .12$; the b values per SOA at $t = 0$ were 0.04, 0, -0.02 and -0.02 . The log factor too showed no significant difference between SOA conditions in a per-time point repeated measures ANOVA with SOA as a single within-subject factor and the subject as a random factor, $F(3, 57) < 1.31$, $p > .28$. In fact, the log factor showed no significant positive contribution in any of the conditions (Figure 8b). Thus, as in Experiment 1, the differences in horizontal movement onset times fully accounted for the log effect as well as for the differences between the four SOA conditions, including the log-linear dissociation.

Factors affecting horizontal movement onset. We next examined how the target number and SOA affect the horizontal movement onset times (Figure 5b). Similarly to Experiment 1, the onset times, specified as the time since the target number appeared

on screen, were analyzed using repeated measures ANOVA with the subject as the random factor and with three within-subject factors: the target side (<20 , left; or >20 , right) and two numeric factors—the SOA and the absolute distance between the target number and 20. To minimize noise, as well as to resolve the problem of missing data in 13 participant-SOA-target combinations, the distance factor grouped each set of three adjacent target numbers, resulting in five levels of this factor: 6–8, 9–11, 12–14, 15–17, and 18–20.

A main effect of SOA, $F(1, 19) = 80.53$, $p < .001$, $\eta_p^2 = .81$, $\eta_c^2 = .24$, mirrored the SOA effect that was earlier observed in the trial-level PRP analysis: Decreasing the SOA created some delay in the movement onset, indicating that the dual task interference was not complete and that there was partial resource sharing with the naming task.

A main effect of side, $F(1, 19) = 17.15$, $p < .001$, $\eta_p^2 = .50$, $\eta_c^2 = .14$, reaffirmed the small-number advantage: As predicted by the differential encoding time model, onset times were earlier for small target numbers (<15) than for large target numbers (>25). We then examined whether the small-number advantage interacted with SOA. The small-number advantage was calculated per SOA as the delta between mean movement onsets on the left and right sides. The differential encoding time model predicts an increasing small-number advantage for smaller SOAs (i.e., for larger overlap between the two tasks). Indeed, averaged over participants, the small-number advantage was 79 ms, 85 ms, 65 ms, and 54 ms for SOA = 0, 100, 200, 300, respectively, and the three-way ANOVA showed that this difference between SOA conditions was significant, Side \times SOA interaction: $t(19) = 1.74$, one-tailed $p = .05$, $\eta_p^2 = .16$, $\eta_c^2 = .004$.

A significant main effect of distance-from-20, $F(1, 19) = 22.86$, $p < .001$, $\eta_p^2 = .58$, $\eta_c^2 = .06$, showed that movement onset was delayed for target numbers closer to the middle of the number line. To examine whether this distance effect was sensitive to the SOA manipulation, we calculated the distance effect per participant and SOA as the slope of the onset-per-target function. This was done using regression analysis with the movement onset time as the dependent variables and with two predictors: the target number side (-1 or 1) and its absolute distance from 20. The resulting b_{distance} from this regression reflects the distance effect; its values for SOAs 0, 100, 200, and 300 were -10.1 ms, -6.6 ms, -5.4 ms, and -4.1 ms, respectively (average over participants), namely, decreasing the SOA continuously increased the distance effect. The three-way ANOVA showed that this effect of SOA on the distance effect was significant, Distance \times SOA interaction: $F(1, 19) = 16.48$, $p < .001$, $\eta_p^2 = .48$, $\eta_c^2 = .01$. There was no significant Distance \times Side interaction, $F(1, 19) = 0.51$, $p = .48$, and no three-way interaction, SOA \times Side \times Distance, $F(1, 19) = 0.13$, $p = .73$.

Discussion of Experiment 2

Experiment 2 used the color naming dual task and manipulated the color-number SOA. The analysis of trajectories replicated the dissociation between the log and linear factors that was observed in Experiment 1: decreasing the SOA decreased (or delayed) the linear factor in the participants' mapping to positions (iEPs), while leaving the log factor almost unchanged. In this respect, the effect of shortening the color-number SOA, a manipulation that presum-

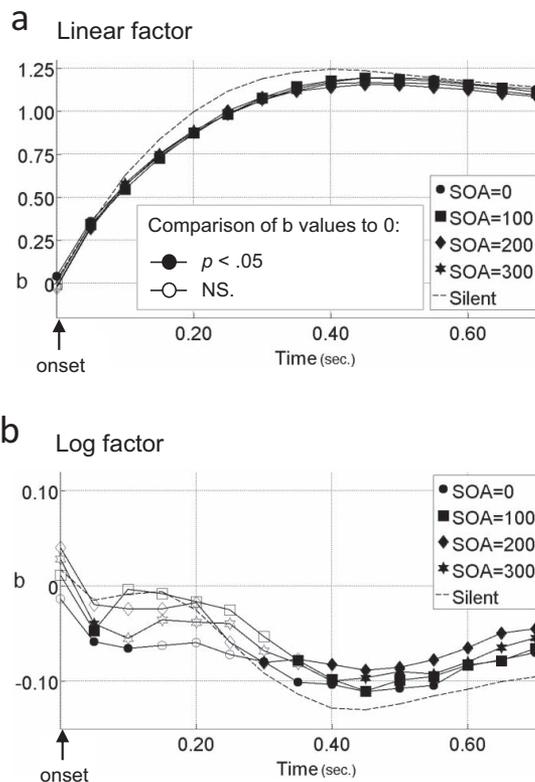


Figure 8. Time course of the effects in Experiment 2 after alignment on horizontal movement onset time. Here, $b[N_{0-40}]$ and $b[\log'(N_{0-40})]$ no longer show any difference between the conditions.

ably makes the experiment harder, was similar to the effect of adding the distracter task in the first place.

The dual representation model can explain these findings as a selective interference of the color naming task with the exact-linear quantity representation, but not with the approximate representation. However, again the differential encoding time model offers a simpler account of the findings. Intertrial differences in the horizontal movement onset times can fully account for the log effect: When the onset times were controlled for (by aligning each trajectory to the trial's movement onset time), the log effect in the regression analyses completely disappeared, and so did the inter-SOA differences in the linear factor.

Experiment 2 also reaffirmed the main assumptions of the differential encoding time model, namely, that horizontal movement onset was earlier for smaller numbers, and that this small-number advantage was increased when increasing the level of interference from color naming (by shortening the color-number SOA).

The use of a PRP design allowed exploring the nature of the interference between the color naming and number-to-position tasks. Several observations in the pattern of delays were compatible with a partial PRP effect. First, *both* tasks were delayed by the interference; in particular, the RT of the color naming task was not constant (as should have been the case if this task was systematically prioritized over the number task), but became slower at shorter SOAs. Second, while the onset of responses to the number task was also delayed at short SOA, the amount of this delay was not compatible with a full PRP effect. The number task was not delayed by a full 100 ms whenever the SOA decreased by this amount, but rather, by about 50%–70% of that value. Third, the size of the target number influenced the horizontal movement onset time of Task 2, but crucially this effect was not additive with SOA (as predicted by a rigid delay of Task 2 due to a full PRP effect; Pashler, 1984, 1994) but was enhanced at short SOAs. All these findings indicate that color naming was not fully prioritized over finger pointing, which is perhaps not surprising given that participants were required to start moving the finger to make the target appear, and were therefore already “launched” in the number-to-position task.

The above observations are compatible with either a partial resource sharing model (Tombu & Jolicoeur, 2002), according to which both decisions are computed in parallel and are jointly slowed by dual-task interference, or by a rigid delay model (Pashler, 1984, 1994; Sigman & Dehaene, 2005) with random prioritization of one task or the other (Sigman & Dehaene, 2006). The latter interpretation predicts that our trials are a mixture of two trials types, depending on whether the central decision does color first and number second, or vice versa. However, given the variability in task performance, this bimodal distribution model cannot be distinguished from the single distribution predicted by partial resource sharing.

Experiment 3: 0–100 Number Line

Experiments 1 and 2 supported the differential encoding time model: Small numbers are encoded faster than large numbers, thereby inducing the transient log effect in the iEPs. The model stipulates that the reason for the small-number advantage is that quantity encoding is noisier for large quantities than for small

quantities (differential variability), and the greater noise causes slower processing. However, an alternative account is that single-digit numbers are processed faster than two-digit numbers—that is, what we observed in Experiments 1 and 2 was not a small-number advantage but a single-digit advantage.

In the setting of Experiments 1 and 2, the two models are hard to tease apart, because over the range of target numbers that were analyzed for movement onset (0–14 and 26–40) most of the small numbers were single digits. To dissociate between the small-number advantage model and the single-digit advantage model, Experiment 3 used a longer number line (0–100), which allows excluding from the analysis the single-digit numbers and consequently the possibility for a confounding factor.

Method

Seventeen right-handed adults (aged $26;10 \pm 5;2$) with no reported cognitive deficits were paid €5 for participation. Their mother tongue was Hebrew. The experiment was performed like the silent condition in Experiment 1, except that the number line extended from 0 to 100 (rather than from 0 to 40). Each number between 0 and 100 was presented four times, that is, 404 nonfailed trials per participant. The horizontal movement onset time was encoded as described in the “Identifying the onset of horizontal movement” section, while excluding trials with target numbers 39–61. Automatic onset encoding succeeded for 82.8% of the trials, and manual encoding increased this to 97.8%. The analyses of onset times (described below) were also performed without the manually encoded trials and the results were essentially the same.

Results

The rate of failed trials in this experiment was $2.9\% \pm 2.3\%$. The mean movement time was $1,191 \pm 204$ ms, the endpoint bias was -0.25 ± 1.13 numerical units, the endpoint error was 4.69 ± 1.87 numerical units, and the horizontal movement onset time was 444 ± 113 ms (all standard deviations refer to the between-subjects variance of the per-subject means). The median trajectories are presented in Figure 9a.

The trajectory data was submitted to regression analysis with iEP as the dependent variable and with five predictors: N_{0-100} , $\log'(N_{0-100})$, the unit digit U, the spatial-reference-points-based bias function SRP, and the target number of the previous trial, $N-1$. One regression was run per participant and time point in 50-ms intervals. The per-subject regression b values of each predictor and time point were compared with zero using a t test. The results (Figure 9b) replicated the previous experiments: dominant linear factor, transient logarithmic factor, SRP contribution in the late trajectory parts, and an effect of the previous trial in early trajectory parts. When the regressions were rerun after aligning each trajectory to the trial's horizontal movement time, the log factor disappeared (Figure 9c), as predicted by the differential encoding time model.

The small-number advantage was found in this experiment too, even when we analyzed only the two-digit numbers (Figure 9d): The horizontal movement onset of targets in the range 10–38 was shorter than that of targets in the range 62–90 by 22 ± 54 ms (the standard deviation refers to the between-subjects variance of the per-subject means). To examine this difference statistically, the

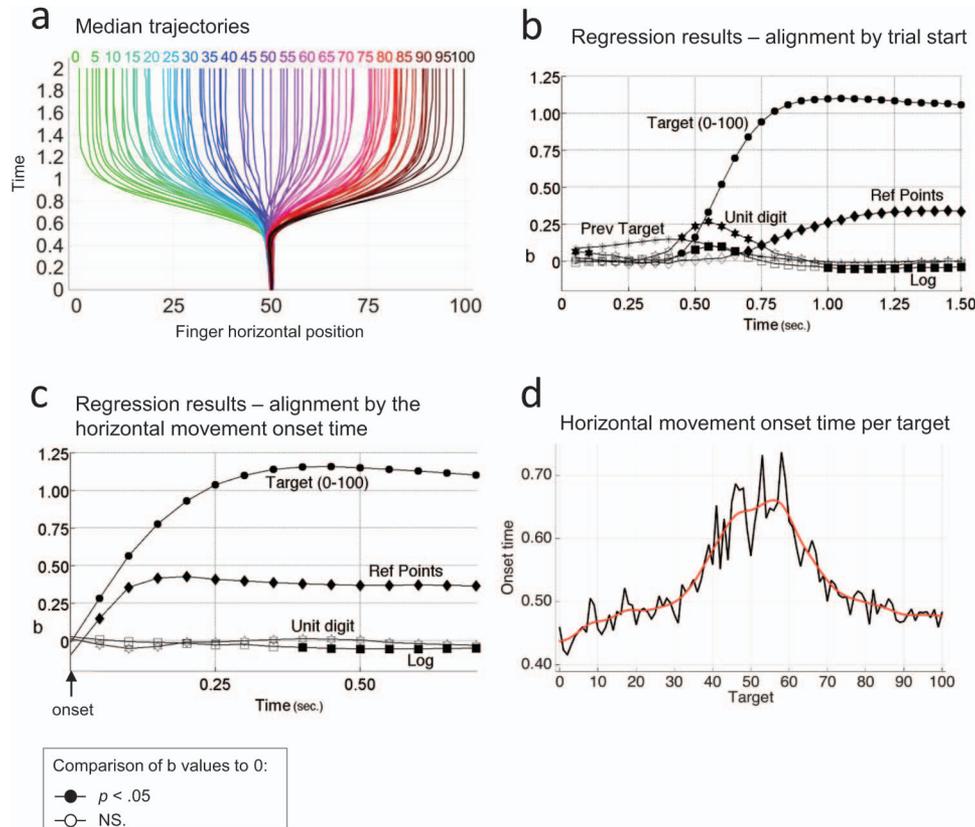


Figure 9. Results of Experiment 3 (0–100 number line). (a) Median trajectories, created by resampling each trajectory into equally spaced time points, finding the per-subject + target median coordinates in each time point, and averaging these medians per target number. (b–c) Regression b values (dependent variable: implied endpoint), averaged over participants. In (b), the trials were aligned to the trial start time and a significant transient log effect appeared. In (c), the trials were aligned to the horizontal movement onset time. This eliminated the log effect, as predicted by the differential onset time model. (d) Mean horizontal movement onset time per target. The black line is the average over trials and participants. The red line is the same data after Gaussian smoothing with $\sigma = 3$. Crucially, a significant small-number advantage was found not only over all targets but also within two-digit numbers, contrary to the notion that it originated only in processing speed differences between single-digit and two-digit numbers. See the online article for the color version of this figure.

onset times were submitted to repeated measures ANOVA with within-subject factors of side (smaller or larger than 50) and distance from middle ($|target - 50|$) and the subject as the random factor. Distance was a numeric factor and it grouped each set of three adjacent targets so the factor had 9 levels (12–14 to 36–38). A significant main effect of side, $t(16) = 1.84$, one-tailed $p = .04$, $\eta_p^2 = .17$, $\eta_G^2 = .04$, confirmed the small-number advantage within two-digit numbers, and thus refuted the “single-digit advantage” hypothesis. The distance effect was also significant, $F(1, 16) = 68.1$, $p < .001$, $\eta_p^2 = .81$, $\eta_G^2 = .18$, with later onset times close to the middle of the number line, and there was no Side \times Distance interaction, $F(1, 16) = 1.84$, $p = .19$. Similar results were obtained when single digits were included in the analysis: Targets 0–39 had shorter movement onsets than 61–100 by 23 ± 55 ms. The Side \times Distance ANOVA showed significant main effects of side, $t(16) = 1.87$, one-tailed $p = .04$, $\eta_p^2 = .18$, $\eta_G^2 = .05$, and distance, $F(1, 16) = 69.2$, $p < .001$, $\eta_p^2 = .81$, $\eta_G^2 = .21$, with no interaction, $F(1, 16) = 0.91$, $p = .35$.

To characterize the distance effect, the horizontal movement onset times were regressed with three predictors: The target side (left = -1 , right = 1), its distance from the middle of the line, and $\log(\text{distance})$, linearly transformed to 0–50. The side effect was significant, $b = -27.78$ ms, $t(6,161) = 6.59$, one-tailed $p < .001$. The $\log(\text{distance})$ effect was significant, $b = -5.28$ ms, $t(6,161) = 9.39$, one-tailed $p < .001$, and much stronger than the linear distance effect, $b = -0.68$ ms, $t(6,161) = 1.75$, one-tailed $p = .04$, in accord with number comparison studies (Cantlon & Brannon, 2006; Dehaene, Dupoux, & Mehler, 1990; Dehaene, 1989).

Discussion of Experiment 3

Experiment 3 showed a small-number advantage, earlier onset of horizontal movement for smaller targets than for large targets, even within two-digit numbers. Thus, the small-number advantage cannot be discarded as faster processing of single digits; it is a genuine phenomenon in processing of two-digit numbers.

Experiment 3 also replicated the other major findings of our previous experiments: The regressions showed a strong linear factor, a transient log factor (which was eliminated when aligning trajectories by the movement onset time), and a spatial-reference-points effect in the late trajectory parts. The replication of these findings using a 0–100 number line confirms that they do not reflect strategies specific for the 0–40 range (e.g., trying to memorize the positions of decade boundaries—a strategy overtly used by several participants in the 0–40 experiments, but not in the 0–100 experiment).

Interestingly, whereas our previous experiments showed that the decade digit was processed parallel to the unit digit (Experiments 1 and 2) or slightly after it (Dotan & Dehaene, 2013), in Experiment 3 the regressions showed a strong effect of the unit digit. Although absent from the aligned-by-onset regressions, this effect suggests decomposed processing of the unit quantity. The exact nature of this decomposed processing cannot be unambiguously determined by the present experiment, and may be a subject for future research (Dotan & Dehaene, 2016; for a discussion of possible interpretations of the unit digit effect, see Dotan & Dehaene, 2013).

Nontransient Logarithmic Effects

The differential encoding time model attributes the logarithmic mapping to delayed horizontal movement onset in trials with large target numbers. Presumably, the effect of this delay will not last forever: Eventually, even the large-target trajectories catch up with the small-target trajectories, and the differences in horizontal movement onset become irrelevant as other factors start governing the finger movement. Thus, the differential encoding time model can account only for a transient logarithmic effect, which disappears in late trajectory parts. Indeed, this was the pattern observed in Experiments 1–3. Several other studies, however, reported nontransient logarithmic effects, which were observed even in the endpoints—in children (Berteletti et al., 2010; Booth & Siegler, 2006; Opfer & Siegler, 2007; Siegler & Booth, 2004) and in a brain-injured adult (Dotan, Friedmann, & Dehaene, 2014).

We hypothesized that the differential encoding time model will not be able to explain such nontransient logarithmic effects. To test this prediction, Experiment 4 examined the number-to-position mapping of fourth-grade children. We also reanalyzed the number-to-position mapping data of ZN, a brain-injured adult who showed a logarithmic effect in the trajectory endpoints (Dotan et al., 2014). We examined whether the log effect in these cases would be observed even when the trajectories are aligned by the movement onset time.

Experiment 4: Fourth Grade Children

Method. Forty-three Hebrew-speaking fourth-grade children (aged $9;9 \pm 0;4$), recruited from a single elementary school in Tel Aviv, Israel, participated voluntarily in this experiment, with written informed consent of their parents. They performed the silent number-to-position mapping task described in Experiment 1. Each number between 0 and 40 was presented four times.

Visual inspection of the results suggested that the children's trajectory data was noisier than the adults'. We therefore calculated several per-participant quality measures and excluded partic-

ipants with especially noisy data. Two measures were based on the finger's initial direction θ_0 . This direction is presumably independent of the target number, and may reflect a bias, noise, or overrelying on prior trials, all of which could potentially disrupt the trajectory analysis. The value of θ_0 was calculated per trial using regression analysis with the x -coordinate as the dependent variable and the y -coordinate as the predictor, over all time points (in 10-ms intervals) from 0 to 160 ms, or from 0 to 100 ms if the first regression was nonsignificant. We excluded one participant whose $\sigma(\theta_0)$ was an outlier (higher than the participants' 75th percentile by at least 150% the interquartile range), and four participants whose mean θ_0 was an outlier to the left or to the right, mean(θ_0) higher than the participants' 75th percentile or lower than their 25th percentile by at least 150% the interquartile range. We also excluded three participants who had low correlation ($r < .6$) between the endpoints and the target number. For the remaining 35 children (aged $9;8 \pm 0;4$), the horizontal movement onset time was encoded per trial as described above ("Identifying the onset of horizontal movement" section), excluding target numbers 15–25. The encoding succeeded for 63.9% of the trials automatically and for 87.3% of the trials after manual encoding.

Results. The median trajectories are presented in Figure 10a. The trajectory data was submitted to regression analysis with the iEP as the dependent variable and with 5 predictors: N_{0-40} , $\log'(N_{0-40})$, the unit digit U, SRP, and the previous target, $N-1$. One regression was run per time point, in 50-ms intervals. These regressions (Figure 10b) showed a strong log effect that lasted until the end of the trial and was observed even in the endpoints (see the endpoints in Figure 10a).

The trajectory data was then submitted to a similar regression in which each trajectory was aligned to the trial's horizontal movement onset time, and the $N-1$ predictor was removed (Figure 10c). This alignment eliminated the log factor from the initial trajectory parts, but a significant log factor was still observed in the late trajectory parts (from 200 ms post movement onset time) and in the endpoints—a finding that is not predicted by the differential encoding time model.

Reanalysis of Patient ZN's Data

ZN was a 73-year-old man who was recovering from a stroke. He was diagnosed with aphasia, severe apraxia of speech, impaired comprehension, dyslexia, dysgraphia, agrammatism, and a selective deficit in converting multidigit numbers to their verbal representation (but not to quantity). In Dotan et al. (2014), we described in detail his performance in several number processing tasks, including the iPad-based number-to-position task, which he performed like the silent condition in Experiment 1, with each number between 0 and 40 being presented four times. To reanalyze ZN's data, we encoded the horizontal movement onset time of each trial using the method described above ("Identifying the onset of horizontal movement" section), excluding target numbers 15–25. This encoding succeeded for 63.3% of the trials automatically and for 95.8% of the trials after manual encoding.

ZN's trajectories are presented in Figure 10d. They were submitted to regression analysis with iEP as the dependent variable and with five predictors: N_{0-40} , $\log'(N_{0-40})$, the unit digit U, SRP, and the previous target, $N-1$. One regression was run per time point, in 50-ms intervals. This regression (Figure 10e) showed a strong log effect that lasted to the end of the trial and was observed even in the endpoints.

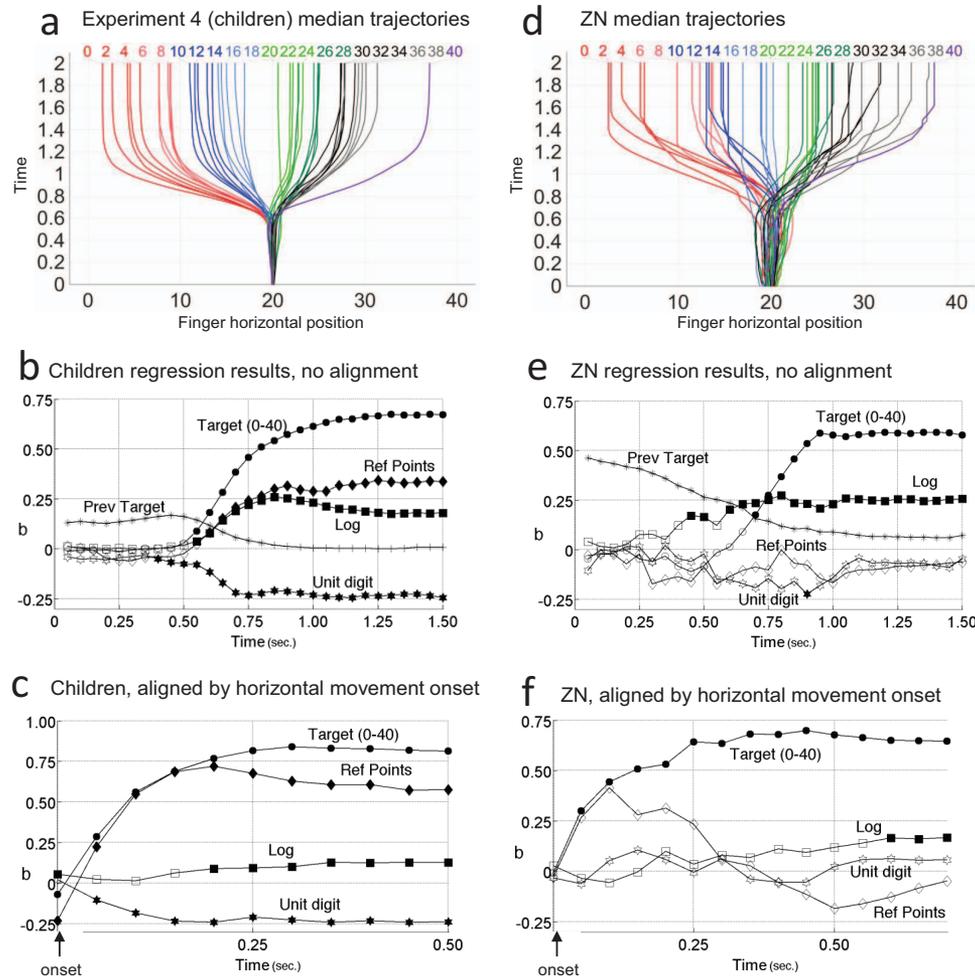


Figure 10. Median trajectories and the regression b values in Experiments 4 (fourth-grade children) and the data of the brain-injured aphasic Patient ZN (a, d) The median trajectories. (b, e) Regression b values, with the trajectories aligned by the target onset. (c, f) The b values of the regression after aligning each trial to its horizontal movement onset time. A significant log effect was found both in Experiment 4 and in ZN's data. This log effect cannot result from different movement onset times per trial, because the alignment by onset controlled for this factor. See the online article for the color version of this figure.

The trajectory data was then submitted to a similar regression in which each trajectory was aligned to the trial's movement onset time, and the $N-1$ predictor was removed (Figure 10f). In line with the differential encoding time model, the log factor was eliminated from the initial trajectory parts. However, contrary to the prediction of the differential encoding time model, a clear log effect was observed in the late trajectory part of the aligned-by-onset regressions (from 600 ms post movement onset time).

Discussion of Experiment 4 and Patient ZN's Data

The main finding from the data of the fourth-grade children (Experiment 4) and of Patient ZN was a nontransient log effect, which was observed in late trajectory parts and in the endpoints. This log effect was not eliminated even when we aligned each trajectory to the trial's horizontal movement onset time. Thus, the log effect cannot be explained by premovement differential pro-

cessing durations, as suggested by the differential encoding time model. We also found no evidence that the log effect in Experiment 4 could be explained by quantity-dependent weighting of prior trials. In this respect, our results were different from Cicchini et al. (2014): Although both studies found logarithmic effect in the endpoints, we did not replicate their finding of larger prior weight for large-target trials. This difference could be related to the fact that we used symbolic targets, while they used a nonsymbolic display (sets of dots).

How should we explain, then, the log effect in the performance of ZN and of the fourth-grade children? We think that two classes of explanations remain tenable. The first class of explanations reverts to the notion of dual quantity representation—linear-exact and approximate. The late log effect would result from amplified approximate representation and decreased exact-linear representation (the *early* log effect may result either from amplified approx-

imate representation or from differential encoding times). The difference between the performance patterns of children and adults in the number-to-position task would then indicate a conceptual log-to-linear shift, as suggested in previous studies (Dehaene et al., 2008; Opfer & Siegler, 2007). Does this log-to-linear shift truly result from a *change* in the quantity representation, which begins as approximate and gradually becomes linear with maturation or education? Or perhaps the log-to-linear shift reflects the addition of a *separate* linear-exact representation on top of the approximate representation, and consistent inhibition of the approximate representation by the exact? The finding of logarithmic mapping in the performance of ZN support the latter possibility: ZN worked as an engineer for many years, and reported being extremely fluent with numbers, so it seems unlikely that his quantity representation remained approximate throughout the years. It also seems unlikely that his brain injury transformed the now-linear quantity representation back into approximate. It seems more likely that his logarithmic mapping reflects an approximate representation that was dormant in his cognitive system and reemerged following a selective impairment to the linear-exact representation.

The second class of explanations for the late log effect is a variant of the differential encoding time model. It assumes that in children and in Patient ZN, unlike in adults, the initial decision to move is based on insufficient evidence. Even if the participants understand the linear requirement of the task and intend to move to the linear position of the target, they may err if the decision process is fed with exceedingly noisy evidence. The participant may then stop short of making the proper inference and start moving based on a partial approximate numerical representation. Because this representation is more precise for small than for large numbers, the movement will be more accurate (more systematically away from the default response) for small than for large numbers, resulting in a log effect. In the General Discussion, we verify this property in a precise mathematical model of the task. In adults, this log bias, if it exists at all, would be quickly compensated by new adjustments of finger position even after the onset of the first horizontal movement, resulting only in a transient log effect. If such a correction is impossible, however, then the log effect will remain sustained.

At present, we cannot decide between those two interpretations. However, the behavioral finding of logarithmic mapping in children is in accord with several previous developmental studies that used number-to-position mapping without tracking trajectories. These previous studies found logarithmic mapping only until second grade (Opfer & Siegler, 2007) or an earlier age (Berteletti et al., 2010; Booth & Siegler, 2006; Siegler & Booth, 2004), whereas here, we found a log effect even in fourth-grade children, that is, in a group that was at least 2 years older. It is possible that our paradigm, which requires a time-limited response and minimal finger velocity, was more demanding than the paradigms used in these previous studies, and therefore increased the logarithmic effect. Such interpretation seems plausible given that, in Experiments 1 and 2, we found that increasing task demands increases the log effect.

A peculiar finding in the children data, which was not observed in any of the adult experiments, is a strong negative effect of the unit predictor in the regressions (Figure 10b, c). In principle, this could mean that the unit effect was either reduced or delayed relatively to the decade effect. However, interpreting this finding

as delayed processing of the unit quantity seems unlikely, because $b[U] < 0$ continues throughout the trial (i.e., the unit digit never catches up with the decade digit). The $b[U] < 0$ can therefore be explained in two ways: Either the decade and unit quantities were not encoded in 1:10 ratio but with underrepresentation of the unit digit, or the unit digit was completely ignored in some trials, resulting in lower $b[U]$ in the regression analysis. Importantly, both explanations suggest that even as late as fourth grade, the processing of two-digit numbers is not fully automated. Previous studies pointed to the log-to-linear shift as one kind of cognitive progress that happens during maturation or education (Berteletti et al., 2010; Dehaene et al., 2008; Opfer & Siegler, 2007); the data from Experiment 4 suggests that assigning proportional weights to the decade and unit quantity may be another cognitive ability that develops with age or education.

Experiment 5: Validating the Movement Onset Detection Algorithm

In all experiments so far, the horizontal movement onset was calculated based on the finger's horizontal velocity profile. To make sure that the onset-detection algorithm did not create some statistical artifact, we administered the number-to-position mapping experiment using a slightly modified paradigm: The participants started moving their finger only after the target number appeared on screen (hereby, stimulus-then-move [StM] paradigm). This is the method used in many trajectory-tracking experiments (e.g., Finkbeiner et al., 2008; Santens et al., 2011; Song & Nakayama, 2008a, 2008b, 2009). While the StM paradigm does not allow for continuous monitoring of cognitive processes at early time points, it has the advantage that the movement onset time can be measured directly rather than calculated statistically.

Method

Twenty right-handed participants aged $28;11 \pm 6;11$ were paid €5 for participation. Their mother tongue was Hebrew and they had no reported cognitive disorders. The method was similar to the silent condition in Experiment 1, except the way a trial was initiated. When the participants touched the initiation rectangle (see Figure 1), a fixation cross appeared, and was replaced by the target number after a random duration between 500 and 1,500 ms. The participants were instructed to move their finger as soon as the target number appeared, but not before that. The movement onset time was registered as the time from stimulus onset until the finger reached the $y = 50$ pixels coordinate (measured from the bottom of the screen). Movement onset lower than 100 ms or higher than 1,000 ms resulted in a failed trial. Each number between 0 and 40 appeared four times, that is, 164 nonfailed trials per participant.

Results

The rate of failed trials was $3.17\% \pm 2.26\%$. The failures were due to moving the finger too early (13.9%) or too late (44%), to violation of the minimal-velocity policy (32%), or to lifting the finger in midtrial (10.1%). The movement onset time was 623 ± 139 ms, and the movement time (from movement onset until reaching the number line) was 529 ± 110 ms. The endpoint bias

was -0.52 ± 0.47 numerical units and the endpoint error was 1.68 ± 0.44 numerical units (all σ refer to the between-subjects variance of the per-subject means).

Figure 11a shows the median trajectories in this experiment. The trajectories are clearly different from the previous experiments: Whereas in the movement-triggers-stimulus (MTS) paradigm the finger initial movement was toward the middle of the number line, in the present experiment the movement was typically aimed more or less directly toward the target number, right from the start. This suggests that the finger movement started only after an initial decision was made about the quantity and the corresponding target position. Note that this pattern is not the result of averaging several trials—it is observed in single trials too (Figure 11b).

The trajectory data was submitted to regression analysis with iEP as the dependent variable and with five predictors: N_{0-40} , $\log'(N_{0-40})$, the unit digit U , the spatial-reference-points-based bias function SRP, and the target number of the previous trial, $N-1$. One regression was run per participant and time point in 50-ms intervals. The per-subject regression b values of each time point and predictor were compared versus zero using a t test. A strong effect of the target number N_{0-40} was found from the time of movement onset (Figure 11c), confirming that the finger aimed more or less toward the target number right from the start. The log effect did not have a significant positive contribution at any time point. There was a clear effect of the SRPs predictor, and there was a $\sim 10\%$ overrepresentation of the unit digit relatively to the decade digit (reflected by the positive contribution of the U predictor). Unlike the previous experiments, no contribution of the previous-target predictor $N-1$ was found at any time point—that is, by the time a decision was made to move the finger, the present-trial quantity has completely overridden the prior trial effect.

The critical analysis in this experiment is that of the movement onset times per target (Figure 11d). The onset times (excluding target = 20) were submitted to repeated measures ANOVA with a between-subjects factor of side (left, right) and a numeric within-subjects factor of distance from 20. A main effect of side, $F(1, 19) = 28.36$, $p < .001$, $\eta_p^2 = .60$, $\eta_c^2 = .02$, confirmed the small-number advantage: Movement onsets for numbers < 20 ($M = 606$ ms) were shorter than for numbers > 20 ($M = 634$ ms). A main effect of distance, $F(1, 19) = 39.74$, $p < .001$, $\eta_p^2 = .68$, $\eta_c^2 = .02$, replicated the findings in previous experiments: Movement onset was shorter when the target number was closer to the ends of the number line. The Side \times Distance interaction was significant, too, $F(1, 19) = 19.03$, $p < .001$, $\eta_p^2 = .50$, $\eta_c^2 = .01$.

A comparison of Figure 11d with Figure 5 shows that the movement onset times in the present experiment (move-then-stimulus-paradigm) were longer than the times detected by our onset-detection algorithm in the MTS experiments. This difference was confirmed by a within-participant analysis: Thirteen of the 20 participants in Experiment 5 also performed the silent 0–40 experiment in the MTS paradigm. The movement onset times of these participants in the StM paradigm (620 ± 84 ms) were longer than the onset times detected in the MTS paradigm (448 ± 46 ms), paired $t(12) = 6.05$, two-tailed $p < .001$, Cohen's $d = 1.68$; targets 15–25 were excluded from this analysis.

Discussion of Experiment 5

The stimulus-then-move (StM) paradigm replicated the major effects found in the movement-triggers-stimulus (MTS) paradigm. In the regression analysis, the finger movement was dominated by the linear quantity representation, with no logarithmic effect—similarly to the aligned-by-movement-onset regressions in Experiments 1–3. This provides further support to the differential encoding time model: When the movement onset is controlled for—either statistically, as in Experiments 1–3, or methodologically, as in the present experiment—the log effect completely vanishes.

The detailed analysis of movement onsets fully replicated the pattern observed in the silent condition in Experiments 1–3: earlier onsets for target numbers on the left side (small-number advantage), and a distance effect such that onset times are later close to the middle of the number line. The replication of these effects with the StM paradigm confirms that these are genuine effects that do not result from a statistical artifact of the onset detection algorithm. This is especially important with respect to the distance effect: The onset detection algorithm relies on the horizontal velocity, and may consequently detect earlier movement onsets when the horizontal velocity is higher, which is typically the case when the target number is closer to any end of the number line. The replication of the distance effect in Experiment 5, in which the movement onset was measured directly rather than calculated, refutes the statistical artifact interpretation and shows that the distance effect has a cognitive origin. Note also that an analogous distance effect was observed by Cicchini et al. (2014, Figure 3b): Their analysis showed higher previous-trial-weights for targets close to the middle of the number line.

How can we explain this distance effect? One possible explanation is inspired by models suggesting that the trigger to change a motor action is the existence of an internal comparison between the action which is intended and the action which is currently being executed (Charles, King, & Dehaene, 2014; Fishbach, Roy, Bastianen, Miller, & Houk, 2007). In Experiments 1–4, participants are asked to initially point toward the middle of the line. Even when the finger is initially at rest (Experiment 5), the motor system might encode a default action of pointing toward the optimal location given the distribution of target numbers, which is again the middle of the number line. As the target-induced intention-to-move builds up, the intention-movement comparison mechanism would predict that the difference between the planned location and the middle of the number line must cross a fixed threshold before the finger starts moving toward the target. What we described in this article as “movement onset” would thus reflect the first decision to change the motor action. The duration of this decision process would be affected by the difference between the default action location (the middle of the number line) and the target number: The farther the target is from the middle of the number line, the larger this difference and therefore, the faster the decision threshold is reached—namely, earlier movement onset time.

Methodologically, Experiment 5 sheds some light on the similarities and differences between the StM paradigm of Experiment 5 and the MTS paradigm of the previous experiments. The StM paradigm may have the advantage of a clearer separation between the two stages involved in this task—the decision stage, whose duration can be directly measured by the movement onset time, and the pointing stage, which is reflected by the finger trajectories.

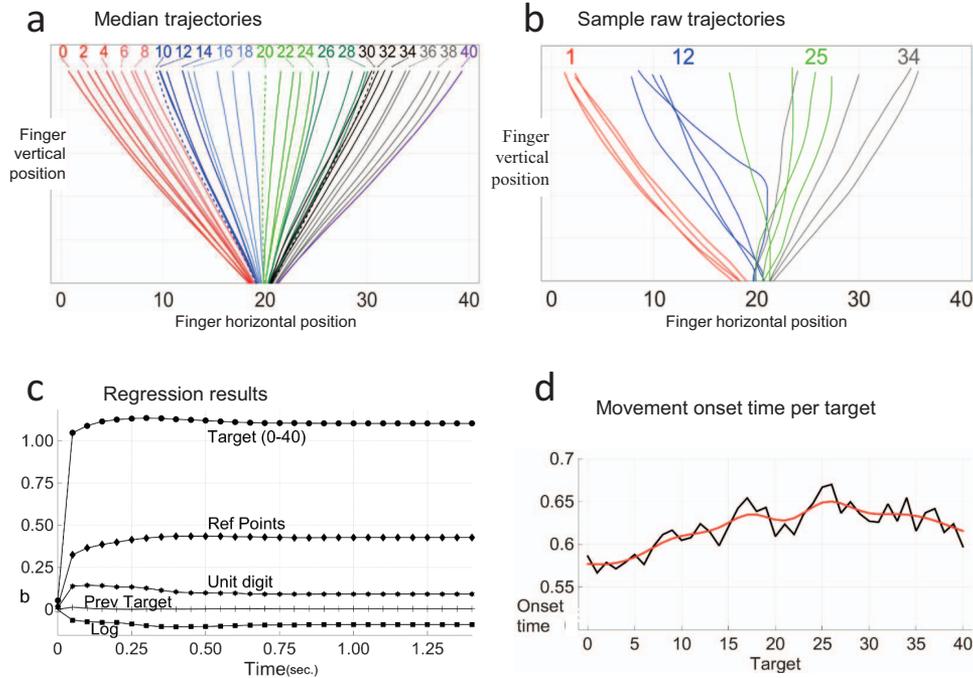


Figure 11. Results of Experiment 5 (stimulus-then-move paradigm). (a) The median trajectories, averaged over participants. (b) Sample raw trajectories of one participant to four specific target numbers. In Panels a–b, the y-axis reflects the iPad screen vertical dimension, so we can see that the finger moves toward the target number right from the start. (c) Regression b values. (d) Movement onset times per target number. The black line is the average over trials and participants. The red line is the same data after Gaussian smoothing with $\sigma = 2$. Onset times were shorter for targets <20 than for targets >20 and were shorter near the ends of the number line than around the middle. See the online article for the color version of this figure.

The StM paradigm also seems to allow for less noisy measurement of movement onsets, as the ANOVAs on movement onset times resulted in much stronger effects in Experiment 5 than in the previous experiments. The MTS paradigm, however, may be superior in its sensitivity to early processes: The onset times we detected in the MTS paradigm were much shorter than the onset times measured in the StM paradigm. One possible reason for this could be that initiating a movement takes longer than changing the direction of an existing movement (Pisella et al., 2000). Another possibility is that the longer onsets in Experiment 5 resulted from the relatively relaxed limit on movement initiation (up to one second from the stimulus onset). Shortening this limit would probably encourage earlier finger movement. Indeed, some implementations of the StM paradigm required participants to initiate movement as quickly as 200–350 ms from the “go” signal (Finkbeiner, Coltheart, & Coltheart, 2014; Finkbeiner & Friedman, 2011). Such short time limits could make the StM paradigm more similar to the MTS paradigm—presumably at the cost of less reliable measurement of movement onset and the duration of the decision stage.

General Discussion

Understanding the Number-to-Position Task

In a series of experiments, we investigated how two-digit Arabic numbers are encoded as quantities in a number-to-position map-

ping task, which forces participants to convert a numeral into a quantity. To analyze the series of stages involved in this task, we obtained a nearly continuous measurement of finger position, and we used a dual-task setting to perturb specific stages. In Experiment 1, the distraction was manipulated by introducing a simultaneous color-naming distracter task and comparing it with the single-task condition. In Experiment 2, we administered only the dual task, and the distraction was manipulated by changing the SOA of the target color and number. An analysis of the finger trajectories showed similar patterns in both experiments: In the experimental conditions with high distraction (color naming in Experiment 1, shorter SOAs in Experiment 2), the participants’ number-to-position mapping became less linear, and in Experiment 1, also more logarithmic—a clear dissociation between the log and linear factors.

A careful analysis of the finger movement, however, showed that this log-linear dissociation cannot be taken as direct evidence for two distinct quantity representations, because a simpler interpretation can account for the results. This interpretation assumes that the finger horizontal movement onset is earlier for smaller target numbers, presumably because their quantity representation is less fuzzy than that of large numbers, which results in faster encoding of small numbers. As a result, the trajectories fan out more quickly for smaller number than for larger numbers, and this induces a transient log effect in the regressions. The interference from color naming further enhances this small-number advantage, thereby increasing the log effect. This interpretation is supported by the finding that the horizontal movement onset time is increased

for larger numbers. As shown by Experiment 3, this small-number advantage cannot be dismissed as a difference between processing single-digit numbers and two-digit numbers. The interpretation is further supported by the finding that aligning the trajectories on movement onset times completely eliminated the logarithmic effect, revealing only a linear mapping of numbers to positions.

Our best interpretation of the data is that the number-to-position mapping task involves separate processes of quantification, decision by evidence accumulation, and pointing (Figure 12a). The quantification process converts the two-digit number into a quantity representation. The decision process maps the quantity representation to a planned position. The pointing process aims the finger to the planned position.

The duration of the decision stage is affected by at least two factors: (a) Number Size: Large numbers take longer to process than small numbers, presumably because of differential variability in the output of the quantification process (in line with previous studies, e.g., Brysbaert, 1995; Li & Cai, 2014; Schwarz & Eiselt, 2009), and (b) Distraction (here induced by the color-naming dual task), which delays the accumulation of evidence arising from the target number (again in line with previous studies of decision making, e.g., Sigman & Dehaene, 2005). Because of partial resource sharing, these two factors interact, so the size of this dual-task delay may also depend on number size, with large quantities suffering from a larger delay than small quantities.

A Mathematical Model of the Number-to-Position Task

To flesh out those ideas, we now present an explicit mathematical model of the number-to-position task. The model provides a “rational” or “ideal observer” analysis, that is, it examines how any rational agent should endeavor to perform this task if it is endowed with exact and/or approximate representations of number. As we will see, such an optimal observer closely predicts human behavior.

We adopt here the same assumptions as in a previous mathematical model of several numerical-decision tasks (Dehaene, 2007). First, at the quantification stage, the quantity associated with the target number is encoded as a time series of independent and identically distributed noisy samples s_t , which are sampled from an internal random distribution. Second, at the decision stage, based on these samples, the posterior distribution over all possible target locations is continuously updated, until a threshold level is achieved and the model commits to a specific location. Third, at the pointing stage, the planned location is used to guide the finger motor movement. We now present detailed equations for each step.

Number representation. Following Dehaene (2007), we assume that within each of the two quantity representation systems, the target number T is represented at any given time step t by a noisy sample $s(t)$ (see Table 4 for a legend of all the notations used here and throughout this mathematical modeling section). The successive samples $s(t)$, $s(t + 1)$, and so forth, are assumed to be independently and identically distributed according to a Gaussian distribution:

$$p(s | T) = \frac{1}{\sigma(T)\sqrt{2\pi}} e^{-\frac{(s-c(T))^2}{2\sigma(T)^2}} = \text{Gaussian}(s, \mu = c(T), \sigma = \sigma(T)) \tag{2}$$

As this expression indicates, the samples s are centered on the value $c(T)$, which is a strictly increasing function of target number T representing the hypothesized internal scale for numerical quantity (e.g., linear or logarithmic). $\sigma(T)$, which may also vary as a function of T , is the standard deviation of the noise on this representation. The choice of functions $c(T)$ and $\sigma(T)$ defines the nature of the internal representation of numbers. For an approximate representation, we may assume either a linear code with scalar variability, that is, $c(T) = T$ and $\sigma(T) = k_1(T + 1)$; or a log-Gaussian coding with fixed variability, that is, $c(T) = \log(T + 1)$ and $\sigma(T) = k_1$. In both cases, k_1 is a constant, and the $+ 1$ term avoids singularity when the target is 0. For an exact representation, we take $c(T) = T$ and $\sigma(T) = k_2$ (where k_2 is another constant).

In the following, we assume, for maximal generality, that exact and approximate representations coexist, are activated in parallel, and generate independent samples. At any time t , the information available for decision is therefore comprised of the two sets of samples from time = 0 to time = t , that is, $\{s_{exact}(t')\}_{t' \leq t}$ and $\{s_{approx}(t')\}_{t' \leq t}$.

Accumulation of evidence. By definition, the ideal observer computes, for every possible response location, the posterior probability that this location is the correct one given the set of past samples. In the number-to-position task, there are as many response locations as there are target numbers, and therefore the inference is equivalent to inferring the likelihood of the current target number being n , given the set of past samples until time = t . Using Bayes’ theorem, we get

$$\text{posterior}_t(n) = p(n | \text{past samples}) \propto p(\{s_{exact}(t')\}_{t' \leq t} | n) p(\{s_{approx}(t')\}_{t' \leq t} | n) p(n) \tag{3}$$

Note that this equation makes uses of the symbol α meaning “proportional to”—this is because, for simplicity, the denominator in Bayes’ rule has been omitted; it is implicitly assumed that the posterior probabilities are normalized by a multiplicative constant to sum to 1 at each time step t .

In Equation 3, $p(n)$ is the prior distribution of target numbers. In the simplest ideal-observer version of the model, the prior is supposed to be flat, in agreement with the fact that, in our experiments, all target numbers in the proposed range are equally likely. Thus, $p(n) = \frac{1}{n_{targets}}$, where n is the number of possible target numbers ($n_{targets} = \text{max} - \text{min} + 1$). Further below, we consider more complex options for the prior.

Given the independence of successive samples, the model reduces to a simple updating rule. Starting from the prior $p(n)$, on each time step t , the optimal observer model receives two new random samples— $s_{exact}(t)$ and $s_{approx}(t)$ —and uses them to update the posterior probability that the correct response is n , using the equation

$$\text{posterior}_t(n) \propto \text{posterior}_{t-1}(n) p(s_{approx}(t) | n) p(s_{exact}(t) | n) \tag{4}$$

(again up to a multiplicative constant, such that the posterior probabilities always sum to 1).

Simulating the random walk inherent to Equation 4 requires expensive computations (generating many trials with random samples at each time step). For a faster, deterministic approximation, we can replace each of the two random multiplicands $p(s_{approx}(t) | n)$ and

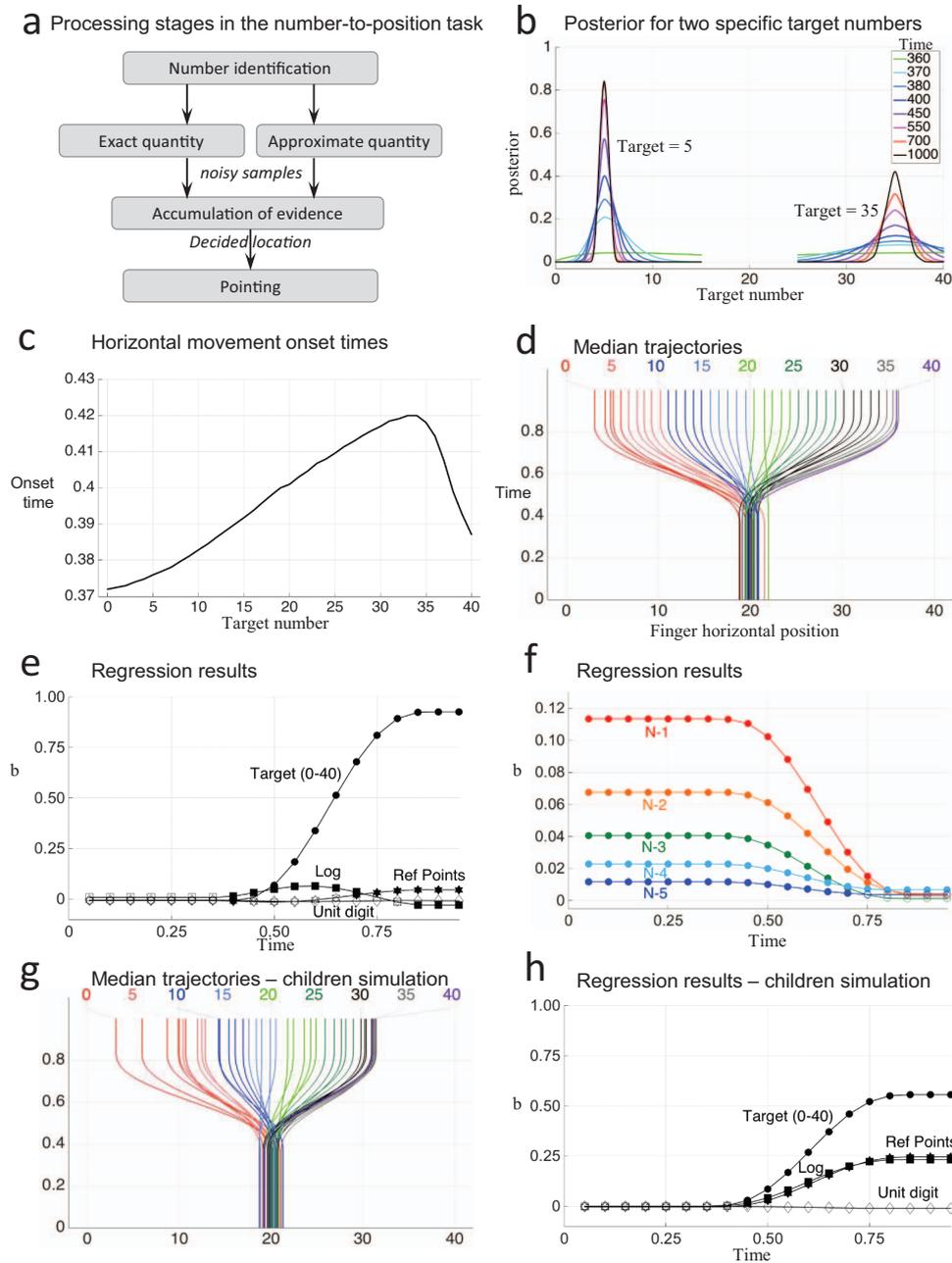


Figure 12. Model and simulations of the processing stages in the number-to-position task. (a) Proposed stages. Incoming digits are identified and the corresponding quantity is separately encoded in approximate and exact systems. Next, evidence accumulation is used to infer the posterior distribution of target locations given the incoming noisy samples. Finally, a pointing stage brings the finger to the location that minimizes pointing errors. (b–f) Simulation results. (b) The posterior probability function in different time points, for two specific target numbers. As the trial progresses the posterior curve becomes steeper. Crucially, the curve converges more quickly for small target numbers such as 5 than for symmetric large target numbers such as 35. (c) Small-number advantage: The horizontal movement onset times are earlier for small target numbers than for larger targets. The onset time were calculated using the onset detection algorithm described in Experiment 1. (d) Median trajectories. (e–f) The regression b values (dependent variable = x -coordinate, predictors = N_{0-40} , $\log'(N_{0-40})$, unit digit, SRP, and the last five targets). The regression captures several effects of the real data—strong linear factor, transient logarithmic factor, and an effect of several prior trials in early trajectory parts, which decays exponentially for older trials. (g) Median trajectories of the simulation of children data. (h) The regression b values of the simulation of children data. See the online article for the color version of this figure.

Table 4
Notations Used for Modeling

Notation	Meaning
T	A target number presented in the experiment
s_{approx}, s_{exact}	A quantity sample sent from the quantification mechanisms (approximate, exact) to the decision process
n	A possible target number (this notation is used mostly for enumeration over all possible targets)
r	A response (decision on a target number) considered by the Bayesian decision process
\hat{r}	The response decided by the participant
λ	The slope of the linear distribution of target numbers, as perceived by the participant. Actual targets were distributed evenly ($\lambda = 0$), but the participants did not know that and may consider various λ values, in distribution denoted $p(\lambda)$.
Gaussian(x, μ, σ)	The probability to get a value x given a Gaussian distribution with mean μ and standard deviation σ
$c(T), \sigma(T)$	The mean and standard deviation of a Gaussian distribution of sample quantities given a target number T
Subscripts	
X_t	The value of X at time point t within a trial
X_i	The value of X at Trial i
Constant parameters in the model	
k_1	Scaling factor for the approximate quantity representation standard deviation
k_2	Standard deviation of the exact quantity representation
k_3	Forgetting: the probability to keep the prior distribution $p(\lambda)$, the perceived target bias ($1-k_3$ is the probability to revert to a flat prior)
θ	Posterior probability threshold for deviating the finger
$\tau_{approx}, \tau_{exact}$	The time (within a trial) in which the quantity samples s_{approx}, s_{exact} start arriving in the decision process

$p(s_{exact}(t) | n)$ by their time-independent expected value. For a trial with target number T , the expected value of s_{exact} is

$$E(p(s_{exact} | n)) = \int_{-\infty}^{\infty} p(s_{exact} | n) p(s_{exact} | T) ds_{exact} \quad (5)$$

The expected value of s_{approx} is calculated with the same formula, replacing s_{exact} by s_{approx} . Although this equation looks symmetrical, note that T represents the target number that was actually presented in the trial, whereas n represents the participant's enumeration of all possible target numbers.

The product of two Gaussians is itself a Gaussian, so Formula 5 yields

$$E(p(s_{exact} | n)) = \text{Gaussian}(c(n), \mu = c(T), \sigma = \sqrt{\sigma(n)^2 + \sigma(T)^2}) \quad (6)$$

and similarly for s_{approx} .

By plugging into this equation the parameters for approximate and exact representation, and by replacing both $p(s_{approx}(t) | n)$ and $p(s_{exact}(t) | n)$ in Equation 4 by their expected values according to Equation 6, we obtain a deterministic approximation of the updating rule for the posterior, given that the target number is T : for log-Gaussian coding,

$$\text{posterior}_t(n | T) \propto \text{posterior}_{t-1}(n | T) \text{Gaussian}(\log(n + 1), \mu = \log(T + 1), \sigma = k_1\sqrt{2}) \text{Gaussian}(n, \mu = T, \sigma = k_2\sqrt{2}), \quad (7)$$

and for linear scalar variability coding,

$$\text{posterior}_t(n | T) \propto \text{posterior}_{t-1}(n | T) \text{Gaussian}(n, \mu = T, \sigma = k_1\sqrt{n^2 + T^2}) \text{Gaussian}(n, \mu = T, \sigma = k_2\sqrt{2}) \quad (8)$$

Numerically, Equations 7 and 8 yield virtually identical results, thus demonstrating the near-complete behavioral equivalence of the log-Gaussian and scalar variability models (Dehaene, 2007). In

the following simulations, we therefore adopt only the log-Gaussian model (Equation 7).

Simulations presented in Figure 12b illustrate how the posterior evolves in the course of the trial for two specific target numbers. Initially, the distribution is flat, and then it evolves to an increasingly sharp peak centered on the target number. Indeed, Equation 7 clearly shows that the ‘‘bump’’ in the posterior distribution is always centered at the appropriate target location on the number line, that is, the highest posterior probability is reached for $n = T$. However, the sharpening of the posterior is faster for small than for large numbers.

Cost function and decision. The above equations specify how the posterior probability distribution of the correct numerical response evolves with time, but not how participants transform this distribution into an intention to move. In any Bayesian decision task, the optimal use of the posterior distribution depends on the cost function imposed by the experimental setting (Maloney & Zhang, 2010). Here, as the task requires minimizing the distance between the finger location and the actual target location on the number line, we stipulate a quadratic cost function:

$$\text{cost}(r) \propto (r - T)^2, \quad (9)$$

where T is the actual target number and r is the subject's intended numerical response. At any time t , we assume that participants pick up, out of all possible responses r , the one that minimizes the expected cost:

$$\hat{r} = \underset{r}{\operatorname{argmin}}(E(\text{cost}(r))) = \underset{r}{\operatorname{argmin}}\left(\sum_n \text{posterior}_t(n | T)(r - n)^2\right) \quad (10)$$

The solution of this equation is the mean of the numbers n , weighted by their posterior probability:

$$\hat{r} = \sum_n \text{posterior}_t(n | T)n. \quad (11)$$

This equation has the following experimental implications: (a) In the absence of information about the target, given that all targets are equiprobable, participants initially point to the center of the number line, that is, the location that minimizes the quadratic error, and (b) as increasingly precise evidence is gathered about the target value, the intended response location deviates progressively from this midpoint value.

Movement. For each target number, the model specifies the subject's optimal intended response at each time step. To compare these numerical estimates with the motor trajectories recorded, we need to model how a numerical intention is translated into a finger trajectory. A complete model would entail answering each of the following theoretical issues: (a) *When* does the finger move? Do participants wait until a threshold amount of evidence is accrued, or is the evidence continuously passed on to the motor system? (b) *How* does the finger move? Is a target direction programmed once per trial, and then translated into a velocity profile? Is the direction updated continuously? Or is it revised only at discrete times, for example, whenever the anticipated finger location deviates from the intended location by a sufficient amount (Fishbach et al., 2007), as suggest by previous "change-of-mind" results (Resulaj, Kiani, Wolpert, & Shadlen, 2009)?

Answering these questions is clearly beyond the present research program. Here, we present simulations of the simplest possible model. Based on prior research on decision making (Gold & Shadlen, 2001), we assume that the decision to move is based on the accumulation of evidence toward a fixed probability threshold θ , that is, movement starts whenever the posterior probability of one of the target locations exceeds this threshold value. At this moment, the movement process sends the finger to the location that minimizes the average square error, as described above. Finally, movement is implemented with the typical bell-shaped velocity profile characterizing limb motion (Flash & Hogan, 1985; Friedman, Brown, & Finkbeiner, 2013).

Under this assumption of a single movement, given that the posterior distribution is sharper for small numbers than for large numbers, the movement onset time should always be slower for large compared to small numbers. However, the choice of the threshold θ has a crucial impact on the shape of the response function. If the participants use a low threshold θ , the finger deviates toward the decided location early on, at a time when the posterior distribution is sharp for small target numbers but not for large ones. This results in a greater separation between small numbers than between large numbers, leading to an approximately logarithmic response pattern (as observed in children and in Patient ZN). If the participants use a higher threshold θ , the finger's deviation toward the decided location happens later, at a time when the posterior distributions for both small and large numbers are already sharply centered on the appropriate target value, so the responses become arrayed in a linear manner.

Effect of prior targets. In our experiments, the prior $p(n)$ was flat over all target numbers. The participants, however, were not told this, and may (explicitly or implicitly) believe that some targets are more likely than others. In agreement with this idea, in all experiments, we observed an effect of the recent target numbers on the early part of the trajectory. Such a prior-trial effect cannot be explained merely as a perseveration of the motor response on the immediately previous trial, because that response was influenced solely by the target of that particular trial and not of the

previous trials. As we shall now see, the exponentially decreasing influence of previous targets can be explained as a constantly updated Bayesian prior over the possible targets.

Formally, we capture this idea using a second-order optimal observer model. The assumption is that subjects use the distribution of recent targets to estimate the probability distribution of a new target T . For simplicity, we assume that participants only consider linear distributions over the range of target numbers, that is, a set of distributions of the form $p(n|\lambda) = \frac{1}{n_{\text{targets}}} \left(1 + \lambda \frac{n - \text{mean}}{\text{max} - \text{mean}} \right)$ with $\text{mean} = \frac{\text{max} + \text{min}}{2}$, where min and max are, respectively, the minimum and the maximum of the range of target numbers. This equation describes a linear probability distribution over the numerical interval $[\text{min}, \text{max}]$. $\lambda \in [-1, 1]$ is a hyperparameter that governs the relative emphasis of small numbers over large numbers: $\lambda = -1$ indicates that participants expect a majority of small numbers, $\lambda = 0$ a flat distribution, and $\lambda = +1$ a majority of large numbers.

We assume that the participants' expectations about the target numbers changes as a function of the recent target numbers they received. This is achieved by constantly maintaining an internal distribution of the possible values of λ . At the beginning of the experiment, this distribution is flat over the interval $[-1, 1]$: All values of λ are equiprobable. At the end of each trial, based on the target they just received, subjects revise their posterior distribution of λ . We denote this revised distribution by $p(\lambda_i)$ (this is the estimate at the end of Trial i , after taking into account the Target T_i , and therefore serving as a prior for Trial $i + 1$). At this time, we assume that the participants have precisely identified the trial's target number T_i , so they can use it to revise their previous distribution $p(\lambda_{i-1})$. According to Bayes' rule, this update should be

$$p(\lambda_i | T_{1:i}) \propto p(\lambda_{i-1} | T_{1:i-1}) p(T_i | \lambda_i) \quad (12)$$

This optimal equation, however, would simply imply that subjects accumulate perfect evidence about the distribution of targets, without any forgetting, in which case they would quickly converge to a distribution centered on the correct value $\lambda = 0$ (unbiased distribution of target numbers). The evidence, however, indicates a strong effect of recent trials, which suggests the existence of local expectations (e.g., after a streak of large numbers, subjects expect to see more large numbers). We model this as forgetting in the updating process. Formally, as in previous work (Behrens, Woolrich, Walton, & Rushworth, 2007; Meyniel, Schlunegger, & Dehaene, 2015), we assume that there is a probability k_3 that the participants carry the current posterior estimates $p(\lambda_{i-1})$ onto the next trial, and a probability of $1 - k_3$ that they revert to a flat prior. In other words, k_3 controls the relative weight of the prior expectation relative to the incoming evidence at a given trial: $k_3 = 1$ means no forgetting (optimal Bayesian integration), and $0 \leq k_3 < 1$ mean underweighting of the prior information and, correspondingly, a stronger effect of the last target on the estimation of λ .

The value of λ_i can now be calculated by applying Bayes' rule:

$$p(\lambda_i | T_{1:i}, k_3) \propto p(\lambda_i, T_{1:i-1}, n_i | k_3) p(T_{1:i}), \quad (13)$$

$$\propto \int_{-1}^{+1} p(\lambda_{i-1}, \lambda_i, T_{1:i-1}, T_i | k_3) d\lambda_{i-1}, \quad (14)$$

$$\propto \int_{-1}^{+1} p(T_{1:i-1} | k_3) p(\lambda_{i-1} | T_{1:i-1}, k_3) p(\lambda_i | \lambda_{i-1}, T_{1:i-1}, k_3) \\ \times p(T_i | \lambda_{i-1}, \lambda_i, T_{1:i-1}, k_3) d\lambda_{i-1}, \quad (15)$$

$$\propto \int_{-1}^{+1} p(\lambda_{i-1} | T_{1:i-1}, k_3) p(\lambda_i | \lambda_{i-1}, k_3) p(T_i | \lambda_i) d\lambda_{i-1}. \quad (16)$$

In Equation 14, we removed the constant term $p(T_{1:i})$ and marginalized over λ_{i-1} . In Equation 15, we applied the chain rule. In Equation 16, we removed the constant term $p(T_{1:i-1} | k_3)$ and simplified the other probabilities by considering that some terms are independent of each other. In the resulting expression (Equation 16), the term $p(\lambda_{i-1} | T_{1:i-1}, k_3)$ reflects the prior; the term $p(\lambda_i | \lambda_{i-1}, k_3)$ – the forgetting factor; and the term $p(T_i | \lambda_i)$ – the probability that the present trial target would indeed be T_i given a certain λ value.

Once we know the distribution $p(\lambda_i)$, we can marginalize over λ_i to obtain the prior probabilities for target number of the next trial:

$$p(n_{i+1} | n_{1:i}) = \int_{-1}^{+1} p(n_{i+1} | \lambda_i) p(\lambda_i) d\lambda_i \quad (17)$$

Intuitively, the effect of those equations is that after receiving, say, a large number such as 40, participants infer that the estimated likelihood of being in an experiment with a large λ is high, and therefore they expect to receive other large target numbers on subsequent trials. As consequence, even in an unbiased experiment where all targets are presented equally frequently, participants will be biased to point toward recently presented targets.

Simulations. Figure 12c–f shows simulations of movement time, movement trajectory and regressor estimates. It can be seen that the model provides a reasonable qualitative fit for most of the experimentally observed effects (here and in Dotan & Dehaene, 2013). The horizontal movement onset is an asymmetrical function of target size, with faster responses for small numbers than for large numbers (Figure 12c). As a result, simulated finger trajectories depart from the center faster for smaller numbers than for larger number (Figure 12d). Consequently, regression analyses exhibit a transient log effect followed by a sustained linear effect (Figure 12e). The log effect disappears when regression is locked on the horizontal movement onset. Finally, an effect of previous targets is observed on the initial part of the movement, with approximately exponential decay over the past trials (Figure 12f).

The model may also account for two additional subtle features of the data: the influence of the SRPs equation, and the fact that the regression weight of the log function becomes negative late in the trial. Both effects arise because the model only considers hypotheses in the range [0,40], thus truncating the posterior distribution to this range and shifting the responses away from the endpoints 0 and 40 and toward the center of the number line (a regression to the mean typical of Bayesian models, see, e.g., Fischer & Whitney, 2014; Jazayeri & Shadlen, 2010). The reference point effect captures this small displacement, while the negative log captures a slight asymmetry of this effect due to differential variability for small and large numbers. In actual data, the reference point effect is larger, seemingly because of an additional repulsion of responses away from the line midpoint 20, which is not captured by the current model (but might be if one assumed an additional process of comparing the target to the midpoint).

The simulations in Figure 12 were obtained with $k_1 = 0.7$, $k_2 = 20.0$, $k_3 = 0.7$, $\theta = 0.15$, with a delay of $\tau_{approx} = \tau_{exact} = 350$ ms for the onset of samples arising from the exact and

approximate representations, and with the assumption of calculation iteration every 1 ms. Because the model remains coarse and unspecified, especially as concerns movement programming, we did not attempt a quantitative fit of the data, but we did observe that the above effects are generic across a larger range of parameters. Scalar and compressive representations of approximate number give virtually identical results. Importantly, having only an exact linear representation cannot account for the results: Simulating it leads to a disappearance of the transient log effect. Conversely, however, it is possible to account for the results with a single approximate representation—there is a range of parameters (e.g., $k_1 = 0.7$, $k_3 = 0.7$, $\theta = 0.12$, $\tau_{approx} = 350$ ms) for which the movement onset is delayed for large numbers, resulting in a transient log, and yet the internal distribution at the time of movement is precise enough to yield near-linear pointing. The only quantitative inadequacy of this approximate-only model is that the weight of the linear regressor never converges to 1, that is, the final pointing remains sublinear. The fact that the linear weight does converge to 1 in adult data (Figure 3, 7, 9b) thus confirms that adults are supplementing their approximate representation with a linear understanding of exact number.

Both the single (approximate) and the dual-representation models can also account for the children’s data by lowering the posterior threshold θ required for making a decision. Lower threshold leads to an earlier decision to move. In this earlier time point, less evidence was accumulated, so the decision about a target location is based on a more approximate representation, thus magnifying the difference between small and large numbers. This results in a more logarithmic mapping (Figure 12g–h, created by lowering the threshold θ from 0.15 to 0.07), which bears much similarity to the children data in Figure 10a–b. Finally, the effect of dual-task interference may be simulated in several ways, either by differing the onset of the exact representation relative to the approximate representation, or by assuming that, during dual-task interference, both representations suffer from additional noise, such that the rate of evidence accumulation is lower. Further research will be needed to disentangle these possibilities.

One aspect that is not captured very accurately by this model is the shape of the distance effect for the movement onset time: The real data show a clear dependency on distance from the midpoint (Figure 5, Figure 11d), which is absent in the simulated data (Figure 12c). This finding suggests that the model’s simple decision mechanism (a fixed threshold on posterior probability, inspired by Gold & Shadlen, 2001) may have to be replaced by a more complex mechanism of comparison between the new aim (point to the target) and the initial aim (point to the midpoint), as indeed suggested by recent studies of motor programming (Fishbach et al., 2007) and error correction (Charles et al., 2014). Such refinements, however, add much complexity to the model and are therefore better left for future research.

Conclusion

Performance of the number-to-position task, as studied in the present experiments with adult participants, is entirely compatible with a strictly sequential processing model that combines a quantification stage (using both exact and approximate representations), an optimal decision-making stage, and a movement stage that minimizes pointing errors. Our main empirical finding is that in

adults, these stages appear to be separable: Once the variable duration of the decision stage is controlled for (by aligning trajectories on the horizontal movement onset times), the finger trajectories show virtually no logarithmic effect, but only linear pointing. Many other details are captured by the optimal decision making model.

While this model nicely accounts for the performance of healthy adult participants, an examination of the performance of the aphasic Patient ZN and of fourth-grade children indicated that this model may not be the whole story. The logarithmic effect in these experiments cannot be solely explained by differential durations of a decision stage, as a logarithmic effect continued to be found long after the horizontal movement onset. We saw that two classes of explanations can be proposed: Either those subjects genuinely fail at the conceptual level, that is, they simply do not understand that the task calls for linear pointing (Booth & Siegler, 2006; Dehaene et al., 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003), or they attempt to point linearly (as our model does), but their decision-to-move is based on partial evidence which is coarser for large than for small numbers. More research will be needed to separate those possibilities.

References

- Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*, 454–459. <http://dx.doi.org/10.1016/j.cognition.2011.11.006>
- Bachoud-Levi, A. C., & Dupoux, E. (2003). An influence of syntactic and semantic variables on word form retrieval. *Cognitive Neuropsychology*, *20*, 163–188. <http://dx.doi.org/10.1080/02643290242000907>
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384. <http://dx.doi.org/10.3758/BF03192707>
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125–135. <http://dx.doi.org/10.1111/j.1467-7687.2010.00962.x>
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221. <http://dx.doi.org/10.1038/nn1954>
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*, 545–551. <http://dx.doi.org/10.1037/a0017887>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*, 189–201. <http://dx.doi.org/10.1037/0012-1649.41.6.189>
- Bormann, T., Seyboth, M., Umarova, R., & Weiller, C. (2015). “I know your name, but not your number”—Patients with verbal short-term memory deficits are impaired in learning sequences of digits. *Neuropsychologia*, *72*, 80–86. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.03.027>
- Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, *124*, 434–452. <http://dx.doi.org/10.1037/0096-3445.124.4.434>
- Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*, 401–406. <http://dx.doi.org/10.1111/j.1467-9280.2006.01719.x>
- Cappelletti, M., Kopelman, M. D., Morton, J., & Butterworth, B. (2005). Dissociations in numerical abilities revealed by progressive cognitive decline in a patient with semantic dementia. *Cognitive Neuropsychology*, *22*, 771–793. <http://dx.doi.org/10.1080/02643290442000293>
- Charles, L., King, J.-R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *The Journal of Neuroscience*, *34*, 1158–1170. <http://dx.doi.org/10.1523/JNEUROSCI.2465-13.2014>
- Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 7867–7872. <http://dx.doi.org/10.1073/pnas.1402785111>
- Cohen, L., & Dehaene, S. (1998). Competition between past and present. Assessment and interpretation of verbal perseverations. *Brain: A Journal of Neurology*, *121*, 1641–1659. <http://dx.doi.org/10.1093/brain/121.9.1641>
- Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology*, *17*, 563–583. <http://dx.doi.org/10.1080/02643290050110656>
- Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology*, *14*, 1029–1061. <http://dx.doi.org/10.1080/026432997381349>
- Dehaene, S. (1989). The psychophysics of numerical comparison: A re-examination of apparently incompatible data. *Perception & Psychophysics*, *45*, 557–566. <http://dx.doi.org/10.3758/BF03208063>
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Attention and Performance, vol. 22: Sensorimotor foundations of higher cognition* (pp. 527–574). Cambridge, MA: Harvard University Press. <http://dx.doi.org/10.1093/acprof:oso/9780199231447.003.0024>
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371–396. <http://dx.doi.org/10.1037/0096-3445.122.3.371>
- Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia*, *29*, 1045–1054. [http://dx.doi.org/10.1016/0028-3932\(91\)90076-K](http://dx.doi.org/10.1016/0028-3932(91)90076-K)
- Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641. <http://dx.doi.org/10.1037/0096-1523.16.3.626>
- 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. <http://dx.doi.org/10.1126/science.1156540>
- Dehaene, S., & Marques, J. F. (2002). Cognitive Euroscience: Scalar variability in price estimation and the cognitive consequences of switching to the Euro. *Human Experimental Psychology*, *55*, 705–731. <http://dx.doi.org/10.1080/02724980244000044>
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974. <http://dx.doi.org/10.1126/science.284.5416.970>
- Dotan, D., & Dehaene, S. (2013). How do we convert a number into a finger trajectory? *Cognition*, *129*, 512–529. <http://dx.doi.org/10.1016/j.cognition.2013.07.007>
- Dotan, D., & Dehaene, S. (2016). *How do we turn a multidigit number into a single quantity?* Manuscript in preparation.
- Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, *63*, 317–351. <http://dx.doi.org/10.1016/j.cortex.2014.08.014>
- Dotan, D., Friedmann, N., & Dehaene, S. (2014). Breaking down number syntax: Spared comprehension of multi-digit numbers in a patient with

- impaired digit-to-word conversion. *Cortex*, 59, 62–73. <http://dx.doi.org/10.1016/j.cortex.2014.07.005>
- Finkbeiner, M., Coltheart, M., & Coltheart, V. (2014). Pointing the way to new constraints on the dynamical claims of computational models. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 172–185. <http://dx.doi.org/10.1037/a0033169>
- Finkbeiner, M., & Friedman, J. (2011). The flexibility of nonconsciously deployed cognitive processes: Evidence from masked congruence priming. *PLoS ONE*, 6, e17095. <http://dx.doi.org/10.1371/journal.pone.0017095>
- Finkbeiner, M., Song, J. H., Nakayama, K., & Caramazza, A. (2008). Engaging the motor system with masked orthographic primes: A kinematic analysis. *Visual Cognition*, 16, 11–22. <http://dx.doi.org/10.1080/13506280701203838>
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 17, 738–743. <http://dx.doi.org/10.1038/nn.3689>
- Fishbach, A., Roy, S. A., Bastianen, C., Miller, L. E., & Houk, J. C. (2007). Deciding when and how to correct a movement: Discrete submovements as a decision making process. *Experimental Brain Research*, 177, 45–63. <http://dx.doi.org/10.1007/s00221-006-0652-y>
- Fitousi, D., & Algom, D. (2006). Size congruity effects with two-digit numbers: Expanding the number line? *Memory & Cognition*, 34, 445–457. <http://dx.doi.org/10.3758/BF03193421>
- Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5, 1688–1703.
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, 2, 59. <http://dx.doi.org/10.3389/fpsyg.2011.00059>
- Friedman, J., Brown, S., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, 57, 140–151. <http://dx.doi.org/10.1016/j.jmp.2013.06.005>
- Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, 5, 10–16. [http://dx.doi.org/10.1016/S1364-6613\(00\)01567-9](http://dx.doi.org/10.1016/S1364-6613(00)01567-9)
- Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, 13, 1020–1026. <http://dx.doi.org/10.1038/nn.2590>
- Kim, D., & Opfer, J. E. (2015). Development of numerosity estimation: A linear to logarithmic shift? *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 1087–1092.
- Li, S. X., & Cai, Y. C. (2014). The effect of numerical magnitude on the perceptual processing speed of a digit. *Journal of Vision*, 14, 1–9. <http://dx.doi.org/10.1167/14.10.1>
- Lourenco, S. F., & Longo, M. R. (2009). Multiple spatial representations of number: Evidence for co-existing compressive and linear scales. *Experimental Brain Research*, 193, 151–156. <http://dx.doi.org/10.1007/s00221-008-1698-9>
- Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*, 50, 2362–2374. <http://dx.doi.org/10.1016/j.visres.2010.09.031>
- Marangolo, P., Nasti, M., & Zorzi, M. (2004). Selective impairment for reading numbers and number words: A single case study. *Neuropsychologia*, 42, 997–1006. <http://dx.doi.org/10.1016/j.neuropsychologia.2004.01.004>
- Marangolo, P., Piras, F., & Fias, W. (2005). “I can write seven but I can’t say it”: A case of domain-specific phonological output deficit for numbers. *Neuropsychologia*, 43, 1177–1188. <http://dx.doi.org/10.1016/j.neuropsychologia.2004.11.001>
- Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS Computational Biology*, 11, e1004305. <http://dx.doi.org/10.1371/journal.pcbi.1004305>
- Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37, 149–157. [http://dx.doi.org/10.1016/S0896-6273\(02\)01144-3](http://dx.doi.org/10.1016/S0896-6273(02)01144-3)
- Núñez, R., Doan, D., & Nikoulina, A. (2011). Squeezing, striking, and vocalizing: Is number representation fundamentally spatial? *Cognition*, 120, 225–235. <http://dx.doi.org/10.1016/j.cognition.2011.05.001>
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8, 434–447. <http://dx.doi.org/10.1037/1082-989X.8.4.434>
- Opfer, J. E., & Siegler, R. S. (2007). Representational change and children’s numerical estimation. *Cognitive Psychology*, 55, 169–195. <http://dx.doi.org/10.1016/j.cogpsych.2006.09.002>
- Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 358–377. <http://dx.doi.org/10.1037/0096-1523.10.3.358>
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116, 220–244. <http://dx.doi.org/10.1037/0033-2909.116.2.220>
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, 44, 547–555. <http://dx.doi.org/10.1016/j.neuron.2004.10.014>
- Pisella, L., Gréa, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., . . . Rossetti, Y. (2000). An “automatic pilot” for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience*, 3, 729–736. <http://dx.doi.org/10.1038/76694>
- Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, 461, 263–266. <http://dx.doi.org/10.1038/nature08275>
- Rouder, J. N., & Geary, D. C. (2014). Children’s cognitive representation of the mathematical number line. *Developmental Science*, 17, 525–536. <http://dx.doi.org/10.1111/desc.12166>
- Ruiz Fernández, S., Rahona, J. J., Hervás, G., Vázquez, C., & Ulrich, R. (2011). Number magnitude determines gaze direction: Spatial-numerical association in a free-choice task. *Cortex*, 47, 617–620. <http://dx.doi.org/10.1016/j.cortex.2010.10.006>
- Santens, S., Goossens, S., & Verguts, T. (2011). Distance in motion: Response trajectories reveal the dynamics of number comparison. *PLoS ONE*, 6, e25429. <http://dx.doi.org/10.1371/journal.pone.0025429>
- Schwarz, W., & Eiselt, A. K. (2009). The perception of temporal order along the mental number line. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 989–1004. <http://dx.doi.org/10.1037/a0013703>
- Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16, 328–331. <http://dx.doi.org/10.3758/PBR.16.2.328>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, 75, 428–444. <http://dx.doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237–243. <http://dx.doi.org/10.1111/1467-9280.02438>
- Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind’s bottleneck. *PLoS Biology*, 3, e37. <http://dx.doi.org/10.1371/journal.pbio.0030037>
- Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual-task and task uncertainty. *PLoS Biology*, 4, e220. <http://dx.doi.org/10.1371/journal.pbio.0040220>

- Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*, 193–208. <http://dx.doi.org/10.1037/a0028560>
- Song, J. H., & Nakayama, K. (2008a). Numeric comparison in a visually-guided manual reaching task. *Cognition*, *106*, 994–1003. <http://dx.doi.org/10.1016/j.cognition.2007.03.014>
- Song, J. H., & Nakayama, K. (2008b). Target selection in visual search as revealed by movement trajectories. *Vision Research*, *48*, 853–861. <http://dx.doi.org/10.1016/j.visres.2007.12.015>
- Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*, 360–366. <http://dx.doi.org/10.1016/j.tics.2009.04.009>
- Tombu, M., & Jolicoeur, P. (2002). All-or-none bottleneck versus capacity sharing accounts of the psychological refractory period phenomenon.

Psychological Research, *66*, 274–286. <http://dx.doi.org/10.1007/s00426-002-0101-x>

Viarouge, A., Hubbard, E. M., Dehaene, S., & Sackur, J. (2010). Number line compression and the illusory perception of random numbers. *Experimental Psychology*, *57*, 446–454. <http://dx.doi.org/10.1027/1618-3169/a000055>

von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: Varieties of developmental dyscalculia. *European Child & Adolescent Psychiatry*, *9*, II41–II57. <http://dx.doi.org/10.1007/s007870070008>

Received May 5, 2015

Revision received June 12, 2016

Accepted June 15, 2016 ■

ORDER FORM

Start my 2017 subscription to *Psychological Review*®
ISSN: 0033-295X

___ \$105.00 **APA MEMBER/AFFILIATE** _____

___ \$250.00 **INDIVIDUAL NONMEMBER** _____

___ \$1,042.00 **INSTITUTION** _____

Sales Tax: 5.75% in DC and 6% in MD and PA _____

TOTAL AMOUNT DUE \$ _____

Subscription orders must be prepaid. Subscriptions are on a calendar year basis only. Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO

American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Call **800-374-2721** or 202-336-5600
Fax **202-336-5568** : TDD/TTY **202-336-6123**
For subscription information,
e-mail: subscriptions@apa.org

Check enclosed (make payable to APA)

Charge my: Visa MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

Billing Address

Street _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

Mail To

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____

REVA17