The perception of number from long-term memory

Jiaying Zhao (jiayingz@princeton.edu)
Department of Psychology, Green Hall,
Princeton University, NJ 08540 USA

Nicholas B. Turk-Browne (ntb@princeton.edu)
Department of Psychology, Green Hall,
Princeton University, NJ 08540 USA

Abstract

The perception of numerosity is supported by two systems: an exact system for small quantities, and an approximate system for large quantities. Two properties arise from the combination of these two systems: the accuracy of numerosity judgments changes qualitatively above the capacity limit for exact representations, and the ability to discriminate two quantities depends on the numerical distance between the quantities and the relationship of this distance to the absolute magnitudes. These well-characterized aspects of number cognition have typically been studied in judgments of numerosity based on visual arrays. Across four experiments we demonstrate remarkably similar effects in numerosity judgments based on incidental long-term memory. These results suggest that similar mechanisms and constraints may operate when estimating numerosity from representations of external sensory input and internal representations derived from long-term memory.

Keywords: Numerosity judgments; perception; memory

Introduction

Perception is typically considered a set of processes for analyzing incoming sensory input. Some researchers have argued that the same perceptual and attentional mechanisms can be directed inward during prospection, memory retrieval, and memory search. To what extent are the mechanisms that underlie judgments on the basis of immediate perception similar or constrained in the same way as the mechanisms that underlie judgments derived from internal representations?

One way to answer this question is to examine the relation between numerical judgments on the basis of immediate external visual input and the numerical judgments based on internal representations from long-term memory. In the former case, several features of immediate numerical perception have been discovered. People are very accurate and fast at enumerating small quantities (6 or fewer), a process termed subitizing (Kaufman, Lord, Reese, & Volkman, 1949), while they are subject to capacity limitations with large quantities, a process termed approximation (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994; Feigenson, Dehaene, & Spelke, 2004). Moreover, when discriminating between numerosities, error rates and response times are inversely related to the numerical distance between numbers (Moyer & Landauer, 1967; Dehaene, Dupoux, & Mehler, 1990). When distance is held constant, error rates and response times increase as the absolute sizes or magnitudes of the two numbers increase (Whalen, Gallistel, & Gelman, 1999; Barth, Kanwisher, & Spelke, 2003). Developmental research has also suggested core systems for representations of exact and approximate quantities (Feigenson et al., 2004; Wood & Spelke, 2005; Opfer & Siegler, 2007).

Numerical judgments from long-term memory have previously been examined in the context of event frequency (Hasher & Zacks, 1979; Hintzman & Block, 1971; Howell, 1973). For instance, according to the strength hypothesis proposed by Hintzman (1969), frequency judgments of an event are determined by the strength or the repetition of the memory trace representing the event. Another view is that frequency judgment is a direct readout of the number of stored traces of an event based on its time lag rather than the strength of a single trace (Hintzman & Block, 1971). More recently, Brown (1995, 1997, 2002) argues that judgments of event frequency depend on context memory in that people rely on enumeration when different contexts produce distinct memory traces.

Here we relate the two areas using tools from studies of immediate numerical perception to focus on how people make numerosity judgments from long-term memory. We investigate the extent to which properties and constraints of numerosity judgments on the basis of long-term memory mirror those of judgments based on immediate perception.

Experiment 1

The purpose of this experiment is to test whether unexpected numerosity judgments from long-term memory are accurate, and whether capacity limitations in subitizing and short-term memory for external visual input also apply for judgments based on retrieved internal representations.

Participants

Twenty students from Princeton University participated in exchange for partial course credit (13 female, mean age 19.2 yrs, *SD* = 1.1).

Materials

Stimuli were chosen from an image set containing 60 distinct object categories. To manipulate numerosity, 50 of these categories were pseudo-randomly assigned to a number between 1 and 10 such that each numerosity level was represented by 5 categories. One exemplar image was chosen from each of these categories, and was presented the corresponding number of times over the course of the first phase of the experiment. For example, if at numerosity level '3' the categories of *dog*, *bear*, *car*, *flower*, and *horse* were chosen, then one exemplar from each category would be presented 3 times throughout the first phase intermixed with images from other numerosity levels. The order of image presentation was randomized for each participant with the constraint that categories could not repeat back-to-back. In addition to the 275 images of interest ($[10+9+8+2+1] \times 5$), 20 additional images were selected randomly from the remaining 10 categories with half presented at the beginning and half at the end to control for primacy and recency effects.

Procedure

In the first phase of the experiment, the participant viewed each image and determined whether it corresponded to a *natural* or *artificial* category by pressing one of two keys. This cover task prevents an explicit strategy such as counting, and is orthogonal to the primary manipulation. On each trial, an image appeared on the screen for 2 seconds, followed by an interstimulus interval of one second. The full trial sequence of 295 images lasted about 20 minutes. The participant then completed an unrelated distractor task for 15 minutes.

In the second phase of the experiment, the participant again viewed single images of objects on the screen, and estimated how many times between 1 and 10 they had seen that image in the first part of the experiment by pressing a number key on the keyboard. The 50 exemplar images of interest were presented in a random order. The accuracy and response time for each image were recorded. Filler images were not presented. It is worth noting that participants often expressed surprise when receiving these instructions, and that in post-experiment debriefing no subject reported being aware that their memory for number would be tested in the second part. These responses suggest that any effects we observe reflect incidental encoding of number in long-term memory.

Results

We compared estimated numerosity from the second phase against the objective numerosity from the first phase. At every numerosity level we averaged across the five categories at that level for each participant, and then averaged these mean estimates across participants. These estimates were compared against the objective numerosity by computing differences within participant and averaging these differences across participants. Results are shown in Figure 1.

To quantify performance, estimated numerosities were modeled as a function of objective numerosities using linear regression. Since estimated and objective numerosities were bounded (from 1 to 10), perfect performance would result in a slope of 1 and an intercept of 0. In contrast, chance performance (i.e. guessing) would lead participants to randomly distribute their responses and would result in a slope of 0. If they randomly distributed their estimates across all response options, the expected intercept would be 5.5 ([10+1]/2). Thus, we can judge accuracy in estimating numerosity from incidental encoding on a continuum from perfect performance (slope = 1, intercept = 0) to chance performance (slope = 0, intercept = 5.5). The linear regression analysis was performed within each participant. The mean

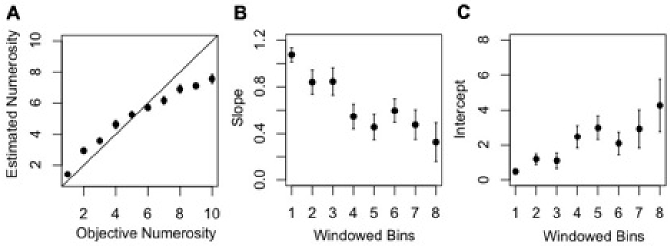

Figure 1: (A) Mean estimated numerosity plotted against the number of times each image was presented during the first phase (objective numerosity). (B) Mean slope of a linear model applied to the data in Figure 1A over windows of three numerosity levels (e.g. '1' reflects window from 1 to 3 on the *x*-axis of Figure 1A). (C) Mean intercept of a linear model applied over the same windows. Error bars reflect std. error.

slope across participants was 0.64 (SD = 0.12, median = 0.64) and the mean intercept was 1.59 (SD = 0.71, median = 1.53).

Prior research has indicated a capacity limitation in highly accurate numerosity judgments of about 4 objects (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Thus, despite overall high performance, accuracy may be non-stationary across objective numerosity. In particular, the slope of linear regressions over smaller windows of objective numerosity may approach 0 (with a corresponding increase in intercept). Such a finding would support the existence of capacity limitation in numerosity judgments from long-term memory.

We thus ran a linear regression across all possible windows of 3 contiguous numerosity levels for each participant. That is, separate linear regressions were run on windows [1,3], [2,4]...[8,10]. For each window, the slope and the intercept values were averaged across participants (see Figure 1B). To quantify our results, one way repeated-measures ANOVAs were performed for slopes and intercepts. There were reliable main effects of numerosity on both measures (slope $F[7,145] = 9.4$, $p < .01$; intercept $F[7, 145] = 3.9$, $p < .01$). Post-hoc Tukey HSD tests revealed that the slope values for window [1,3] (M = 1.08, SD = 0.28) were reliably higher than the rest of the slope values, while the intercept values for the same window (M = 0.50, SD = 0.65) were reliably lower than the rest of the intercept values.

These results suggest that performance starts off near perfect, and declines steadily as a function of objective numerosity. From inspection of Figure 1B, there appears to be a marked drop in slope and increase in intercept after window [3,5], suggesting a capacity limitation around 4-5 repetitions. To quantify these intuitions, we imposed a mixture of perfect and chance performance on the data in Figure 1A. In particular, we tested a mixed linear model to identify at what point along the objective numerosity line at which participants' performance started to level off and to decline. In this mixed model, at a given point n on the numerosity line, $y = 1 \times x + 0$ for x in $[0, n]$, and $y = 0 \times x + 5.5$ for x in $[n+1, 10]$. In other words, we fit the perfect performance linear model to data up to numerosity n and a chance performance model to data from numerosity n + 1 to 10. It should be noted that at numerosity 0 the mixed model becomes a complete chance model and at numerosity 10 it becomes a complete perfect performance model. The average model fits across participants are shown in Figure 2A. SS*error* was minimized at n from 4 to 8.

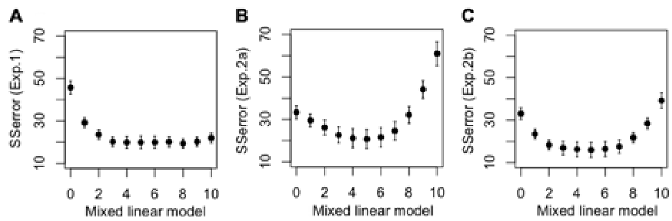

Figure 2: Estimated numerosities in (A) Exp. 1, (B) Exp. 2a, (C) Exp. 2b, tested against the following model: at a given point n on the x-axis, $y = 1 \times x + 0$ for x in [0, n], and $y = 0 \times x + 5.5$ for x in $[n+1, 10]$. Error bars reflect std. error.

Discussion

The results in Exp. 1 demonstrate that participants can make remarkably accurate numerosity judgments from long-term memory. Moreover, participants were unaware that their memory for number would be tested, and therefore this accuracy reflects incidental/automatic encoding of number in memory. Despite overall high accuracy, estimates of numerosity became less accurate when an image has been presented more than 5 times. To our knowledge, this provides the first demonstration of capacity limitations in judgments operating over internal representations, extending and replicating robust findings of similar limits in short-term memory for external visual input (Xu & Chun, 2006).

Experiment 2a

This putative capacity limit observed in Exp. 1 could reflect the fact that we repeated identical images many times. Such repetition could lead to habituation or reduced attention that would impair further encoding. To test this explanation, here we replicate Exp. 1, but present multiple exemplars of the same category once, rather than the same exemplar multiple times. This increased novelty may improve encoding and may facilitate retrieval.

Participants

Twenty students from Princeton University participated in exchange for partial course credit (14 female, mean age 19.7 yrs, $SD = 1.5$). None had served in the previous experiment.

Materials

The materials were identical to Exp. 1 with one important exception: instead of presenting the same exemplar image from each category n times, n distinct exemplars were randomly drawn from each category and presented only once. For example, if the category *dog* was assigned to the numerosity level '3', then images of three different dog breeds would each be presented once. This increased the novelty and variance within each category, possibly allowing for more accurate estimates about large numerosity levels.

Procedure

The procedure was identical to Exp. 1 except for one aspect of the second phase: category names (e.g., "dog") were used to elicit estimates of how many images of that category had been presented in the first phase. Names were used rather than images because there were several possible images to choose from for many of the categories. Thus, 50 category names were presented in a random order.

Results

Data were analyzed in the same manner as Exp. 1. Results are shown in Figure 3. To quantify performance, we modeled estimated numerosities as a function of objective numerosities using linear regression (as in Exp. 1). Surprisingly, the mean slope across participants was 0.33 ($SD = 0.16$, median = 0.34), reliably lower than the mean slope ($M = 0.64$) in Exp. 1 ($t[38] = 6.9$, $p < .01$). The mean intercept across participants was 2.72 (SD = 1.18, median = 2.58), reliably larger than the mean intercept (M = 1.59) in Exp. 1 ($t[38] = 3.7$, $p < .01$). Contrary to our predictions, these results suggest that performance was worse in Exp.t 2a vs. Exp. 1, i.e. farther from perfect performance, and closer to a chance uniform distribution.

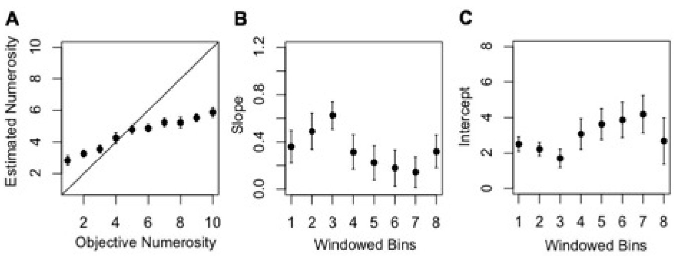

Figure 3: (A) Mean estimated numerosity plotted against the number of exemplars of each category from the first phase. (B) Mean slope of a linear model applied to the data in Figure 3A over windows of three numerosity levels. (C) Mean intercept of a linear model applied over the same windows. Error bars reflect std. error.

We again explored the presence of a capacity limit by computing the slopes and intercepts of linear functions over windows of objective numerosity. Despite the relatively poorer performance in this experiment, visual inspection of Figures 3B and 3C revealed a qualitative difference between windows [3,5] and [4,6]. One way repeated-measures ANOVAs revealed a main effect of numerosity on intercept values ($F[7, 145] = 5.6$, $p < .01$). The main effect of numerosity on slope values did not reach significance ($F[7, 145] = 1.5$, $p > .05$).

To further characterize a change in accuracy as a function of numerosity we also tested the same mixed linear models on the data in Figure 3A. The model fits are shown in Figure 2B. Again, across numerosities the data were best represented by a model in which performance was perfect up to 5 exemplars and plateaued for larger numbers.

Discussion

Contrary to our predictions, providing multiple exemplars for number estimation did not improve accuracy. In fact, performance was worse than in Exp. 1, where judgments were based on the number of repetitions of a single stimulus. This suggests that performance in Exp. 1 did not asymptote around 5 presentations because of habituation or diminished attention. The worse performance here could reflect poor encoding of images presented only once, or source confusion during retrieval in response to a category label. For example, "dog" may retrieve more than 3 exemplars, with reduced performance reflecting an inability to distinguish exemplars intruding from prior experience. Regardless, despite overall worse performance, participants nevertheless showed consistent capacity limitations to Exp. 1 of approximately 5 memories.

Experiment 2b

While the worse performance in Exp. 2a vs. Exp. 1 can be due to weaker encoding of number, it remains possible that potentially more accurate judgments were hampered by a less informative retrieval cue. To examine this possibility, here we replicate Exp. 1 with category labels during retrieval.

Participants

Twenty students from Princeton University participated in exchange for partial course credit (12 female, mean age 19.1 yrs, *SD* = 1.3). None had served in previous experiments.

Materials

The stimuli used here were the exactly same as those in Exp. 1. The exemplar image from each category was repeatedly presented depending on the numerosity value.

Procedure

The procedure was identical to Exp. 2a. Participants were cued by a category name (e.g. "dog") and estimated how many times they had seen an image from that category.

Results

Results are shown in Figure 4. We modeled estimated numerosities as a function of objective numerosities using linear regression. The mean slope across participants was 0.46 (*SD* = 0.15, median = 0.48), reliably higher than the mean slope in Exp. 2a (M = 0.33; $t[38]$ = 2.6, $p < .01$) but reliably lower than that in Exp. 1 (M = 0.64; $t[38]$ = 4.2, $p < .01$). The mean intercept across participants was 2.31 (*SD* = 0.82, median = 2.06), which was not statistically smaller than the mean intercept in Exp. 2a (M = 2.72; $t[38]$ = 1.3, $p > .05$) but was reliably larger than that in Exp. 1 (M = 1.59; $t[38]$ = 2.9, $p < .01$). Thus, after matching retrieval cues, numerosity judgments from multiple repetitions of the same exemplar remain more accurate than those from multiple exemplars of the same category.

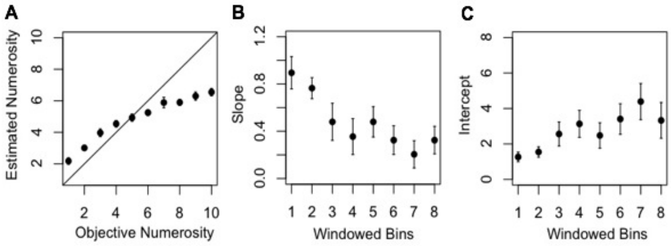

Figure 4: (A) Mean estimated numerosity plotted against the number of exemplars of each category from the first phase. (B) Mean slope of a linear model applied to the data in Figure 4A over windows of three numerosity levels. (D) Mean intercept of a linear model applied over the same windows. Error bars reflect std. error.

To explore possible capacity limitations, we computed slopes and intercepts for linear models over windows of three contiguous numerosities. One way repeated-measures ANOVAs revealed main effects of numerosity on both measures (slope $F[7, 145]$ = 3.2, $p < .01$; intercept $F[7, 145]$ = 4.6, $p < .01$). Post-hoc Tukey HSD tests revealed that the slope value for window [1,3] (M = 0.90, *SD* = 0.61) were reliably higher than the rest of the slope values, while the intercept for the same window (M = 1.26, *SD* = 1.20) were reliably lower than the rest of the intercepts.

Moreover, from Fig. 1, 3, and 4 the slope and intercept values for Exp. 2b appear to resemble those in Exp. 1 more than those in Exp. 2a. Collapsing across participants, the mean slope values were highly correlated with those in Exp. 1 (r = 0.84, $p < .01$), but not with those in Exp. 2a (r = 0.51, $p > .05$). Capacity limitations were examined for each numerosity level where performance plateaued (see Fig. 2C). As in previous experiments, the mixed model fit best at 5.

Discussion

This experiment reveals that judgments of numerosity are more accurate for multiple repetitions of the same exemplar than for single presentations of multiple exemplars of the same category. The category label did somewhat impair performance, but critically, cannot entirely explain the poor performance in Exp.2a. Across three experiments we observed evidence that unexpected judgments of numerosity from past experience are accurate and subject to capacity limitations.

Experiment 3

Capacity limitations are a signature property of perceptual number processing. To further build our case for a similarity between external and internal number estimation, we consider two additional classic effects in the numerosity literature: the *distance effect* and the *magnitude effect*. These psychophysical effects are evident when two quantities must be discriminated. The distance effect refers to the relative ease with which participants can discriminate two quantities that are farther apart in number space (e.g., 2 vs. 3 compared to 2 vs. 4). The magnitude effect refers to the fact that a given numerical distance can become harder to discriminate at higher magnitudes (e.g., 2 vs. 3 compared to 8 vs. 9). We explore whether these psychophysical effects also occur when discriminating numerosities defined by long-term memory.

Participants

Twenty students from Princeton University participated in exchange for partial course credit (11 female, mean age 19.6 yrs, $SD = 1.6$). None had served in previous experiments.

Materials

The stimuli in Exp. 1 were used. The exemplar image of a category was repeatedly presented depending on the numerosity value.

Procedure

The first phase was identical to Exp. 1, such that all number encoding was incidental. In the second phase, participants judged which of two images they had seen more times during the first phase. Based on numerosity levels, we paired images so as to fully cover the space of possible distances and proportional distances (for the magnitude effect). At distance of 1 we paired an image that was presented n number of times with the one that was presented $n + 1$ number of times. Since numerosity levels range from 1 to 10, there were 9 pairs at the distance of 1 (e.g., 1 vs. 2, 2 vs. 3, etc.). The same pairing method was applied to the distances of 2, 3, 4, and 5, which resulted in 8, 7, 6, and 5 pairs, respectively. Thus, a total of 35 pairs were generated. For each pair, two images were presented side by side on the screen and participants judged which image appeared more times by pressing one of two buttons for left or right. The order of pairs was randomized for each participant and the position on the screen of the image with the larger numerosity was randomized on each trial.

Results

To assess distance effects, we pooled all of the pairs of each distance within participant and computed mean accuracy and RT. To assess magnitude effects, we conditioned every distance on the smaller number of the pair (e.g., 1 vs. 3: distance = 2/base = 1). Mean accuracy and RT were again computed within participant for each bin. Results are shown in Fig. 5.

As visible in Fig. 5A and 5B, accuracy increased while response time decreased as a function of distance (the opposite of a speed/accuracy tradeoff). Collapsing across participants, there was a strong correlation between accuracy and distance ($r = 0.98$, $p < .01$) but a weaker correlation between response time and distance ($r = -0.61$, $p > .05$). To test the reliability of these relationships, we correlated accuracy and response time with distance within each participant, transformed the resulting correlation coefficients to Z scores using Fisher's

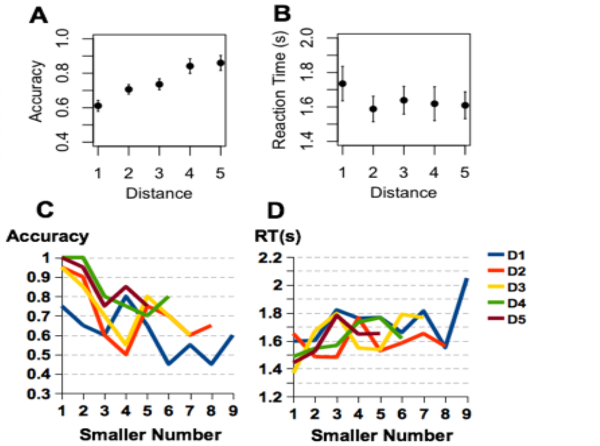

Figure 5: (A) Mean accuracy as a function of distance. (B) Mean response time as a function of distance. (C) Mean accuracy as a function of the smaller number in a pair and also the distance (e.g. D1 = distance of 1). (D) Mean response time as a function of the smaller number and also the distance. Error bars reflect std. error.

transform, and then compared these values against the null relationship of 0 using a one-sample t-test. Accuracy was positively correlated with distance (mean $z = 0.21$, $t(19) = 5.72$, $p < .01$), and response time was negatively correlated with distance (mean $z = -0.08$, $t(19) = 2.15$, $p < .05$). Thus, participants were able to discriminate between two numerosities in long-term memory, and performed better as a function of the absolute number difference.

To assess the magnitude effect, we performed repeated-measures ANOVAs on accuracy and response time as a function of the base and the distance of each pair. We could not include both the base and distance factors in a two-factor ANOVA because the design was not factorial (e.g., there were no distance = 5/base = 9 trials), and thus used separate one-way ANOVAs. There were main effects of base and distance on accuracy (base $F[8, 683] = 18.5$, $p < .01$; distance $F[4, 691] = 50.1$, $p < .01$). There were also main effects of base and distance on response time (base $F[8, 683] = 16.4$, $p < .01$; distance $F[4, 691] = 13.9$, $p < .01$). The robust effect of base demonstrates a magnitude effect in numerosity judgments from memory.

Discussion

When judging which exemplar appeared more times, participants were more accurate and faster when the distance was larger. Holding the distance constant, participants were also more accurate and faster when the base number of presentations was relatively small.

General Discussion

We have found that unexpected numerosity judgments based on long-term memory can be highly accurate, and that this accuracy is maintained for up to a small quantity of retrieved memories. These findings are largely in agreement with studies of numerical judgments based on immediate visual perception. Moreover, judgments of numerosity were more accurate for multiple repetitions of the same exemplar than for single presentations of multiple exemplars. This rules out the possibility that apparent capacity limitations reflect habituation or reduced attention. The fact that multiple exemplars were in fact *worse* than single exemplars could relate to failures of source monitoring (Dougherty & Franco-Watkins, 2003; Johnson, Hashtroudi, & Lindsay, 1993): when cued by a category, participants may have been unable to screen out extra-experimental memories. Such failures may have been minimized when retrieval was cued by an image rather than a category label, but the effect persisted when retrieval cue was equated. When discriminating between two incidentally encoded numerosities, performance increased with distance but decreased as the magnitude or the absolute size of the numbers increased. These results were again in line with findings on numerosity comparison based on immediate visual perception (Whalen et al., 1999; Barth et al., 2003).

Since numerosity judgments based of long-term memory exhibited similar properties and constraints as compared to immediate perception, our findings are consistent with the existence of a common underlying numerosity mechanism for perception and memory. While our focus has been on drawing this analogy, there may also be important differences between the perception and memory of number information. For example, while the perception of number has been well-characterized by a hard capacity limit on exact judgments, memory for number may be better characterized by a more continuous logarithmic or power law function. Moreover, we do not yet know whether input representations retrieved from long-term memory are the same as those constructed during online perception (e.g., whether numerosity is estimated over a set of retrieved episodes, or directly read out from a symbolic or analog representation of quantity updated during encoding). These are important questions for future research, but our results nevertheless provide initial evidence for a striking symmetry between snap judgments of number from a single sensory stimulus, and delayed (surprise) judgments based purely on long-term memory. In sum, mechanisms that seem to exist in the service of sensory processing may have broader functional roles in cognition, operating similarly over input from internal or external sources.

References

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1539–1553.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 898–914.

Brown, N. R. (2002). Frequency processing and cognition. In P. Sedlmeier & T. Betsch (Eds.), (chap. Encoding, representing, and estimating event frequencies: A multiple strategy perspective). Oxford: Oxford University Press.

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*, 625–641.

Dougherty, M. R. P., & Franco-Watkins, A. M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica*, *113*, 23–44.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*, 307–314.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Epxerimental Psychology: General*, *108*, 356–388.

Hintzman, D. L. (1969). Apparent frequency as a function of frequency and the spacing of information. *Journal of Experimental Psychology*, *80*, 139–145.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*, 297–306.

Howell, W. C. (1973). Representation of frequency in memory. *Psychological Bulletin*, *80*, 317–331.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*, 498–525.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1–22.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519–1520.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology*, *55*, 169–195.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80–102.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, *10*, 130–137.

Wood, J. N., & Spelke, E. S. (2005). Chronometric studies of numerical cognition in five-month-old infants. *Cognition*, *97*, 23–39.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.