Human nonverbal discrimination of relative and absolute number

Lavinia Tan · Randolph C. Grace

Abstract The nonverbal discrimination of relative and absolute number of sequential visual stimuli was investigated with humans in bisection, reproduction, and report tasks. Participants viewed a sequence of 40 red and black objects on each trial, randomly intermixed, and had to identify the number of red objects, which varied from 1 to 20. To prevent the use of a verbal-counting strategy, participants were required to name the objects as they appeared. The characteristics of human performance resembled those of pigeons in analogous procedures (Tan & Grace Learning and Behavior 38:408-417, 2010; Tan, Grace, Holland, & McLean Journal of Experimental Psychology 33:409-427, 2007): Average response number increased systematically with sample number, and bisection points were located at the arithmetic, not the geometric, mean. Additionally, in both the reproduction and report tasks, coefficients of variation decreased for values less than 6 but increased or remained constant for larger values, suggesting that different representations were used for small and large numbers.

Keywords Numerical processing · Nonverbal · Number representation · Numerosity

Numerical ability has been widely investigated in humans, often using adaptations of nonhuman numerical discrimi-

L. Tan (✉)
Reed College, Department of Psychology,
Portland, OR 97202, USA
e-mail: ltan@reed.edu

R. C. Grace (✉)
University of Canterbury, Department of Psychology,,
Christchurch, New Zealand
e-mail: randolph.grace@canterbury.ac.nz

nation procedures. Yet relatively few studies have directly compared human and nonhuman performance, and those have been limited mostly to nonhuman primates and relative numerosity discriminations (Beran, Johnson-Pynn, & Ready, 2008; Jordan & Brannon, 2006a, b). The aim of the present experiment was to investigate further the processes underlying the nonverbal discrimination of number; it replicates and extends previous research with pigeons by Tan, Grace, Holland, and McLean (2007) and Tan and Grace (2010) with human participants in both relative and absolute numerical tasks.

Relative numerosity discriminations

One procedure commonly used to investigate relative numerosity discrimination in various species is the bisection task. In Meck and Church's (1983) original experiment, rats were trained to discriminate between two sequences of white noise bursts in which duration and number were confounded; responses to the left or right lever were reinforced depending on whether the preceding sequence was small/short or large/long. Numerical and temporal discriminations were tested, respectively, by holding duration or number constant at an intermediate value while varying the other on test trials. For both discriminations, bisection points were located near the geometric mean of the extreme values used in training.

The location of the bisection point has implications for the representation of number. Assuming similar representational structures for time and number, arithmetic mean bisection would characterize a linear numerical scale, with constant spacing and generalization between values. Conversely, bisection at the geometric mean is predicted by either a logarithmic scale with decreasing spacing and constant generalization between values or a linear scale

with constant spacing and increasing generalization between values (see Gibbon, 1981). Similar to Meck and Church (1983), studies with nonhumans have typically obtained bisection at the geometric mean (Beran et al., 2008; Breukelaar & Dalrymple-Alford, 1998; Emmerton & Renner, 2006; Fetterman, 1993; Jordan & Brannon, 2006a, b; Roberts, 2005; Roberts & Mitchell, 1994).

Numerical bisection has also been investigated in adults (Roitman, Brannon, Andrews, & Platt, 2007) and children (Beran et al., 2008; Droit-Volet, Clement, & Fayol, 2003; Jordan & Brannon, 2006a). In these experiments, temporal and visual nonnumerical cues were varied to limit confounds, and procedural manipulations, such as response time limits and verbal distractor tasks, prevented verbal counting. For both adults and children, results were similar to those obtained with nonhumans: The proportion of *large* responses increased with number, and bisection points were located at the geometric mean, consistent with Weber's law and supporting a representation of number that produces scalar variability. Additionally, bisection functions of children and primates could be superimposed when plotted along the same scale (Beran et al., 2008; Jordan & Brannon, 2006a).

However, Droit-Volet et al. (2003) have presented an exception to the nonverbal bisection data. Their study examined the role of number and time in a bisection task with children and adults. Participants in a counting group were instructed to count aloud at a comfortable rhythm, while those in a noncounting group had a distractor task of saying repetitive speech as rapidly as possible to suppress verbal counting. In this study, bisection points were similar for both time and number and were near the arithmetic mean. There is no clear explanation for why Droit-Volet et al. found arithmetic mean bisection, whereas other studies have generally found nonverbal bisection at the geometric mean. It is possible that the ratio of the largest and smallest values or the spacing of the numerical values influences bisection point locations (Wearden & Ferrara, 1995). Overall, the factors that determine the location of bisection points and the implied nature of numerical representations used by humans and nonhumans in these tasks remain unclear.

Absolute numerical discriminations

Do human and nonhuman discriminations of absolute number exhibit similar response characteristics? Response variability patterns provide some indication of numerical processes and numerical representation in absolute discrimination tasks. Nonverbal numerical discrimination is characterized by scalar variability and conforms to Weber's law: Coefficients of variation (CVs), a measure of relative response variability calculated as the standard deviation of response number divided by mean response number, should remain constant as magnitude increases when plotted on a log-log scale (Cordes, Gelman, Gallistel, & Whalen, 2001). This is also consistent with either a numerical representation that is logarithmic with constant generalization or a linear numerical representation with increasing generalization. Conversely, with verbal counting and a precise numerical representation that is linear with constant generalization, variability should be limited to miscounts—for example, counting an item twice or skipping an item, which are equally likely with every count. Such errors produce binomial response variability, which increases proportionally to the square root of the target numbers. Binomial variability results in decreasing CVs as a function of number, with a slope of –0.5 when plotted on a log-log scale (Cordes et al., 2001).

A separate mechanism for the discrimination of small numbers (less than four or five) has also been proposed. *Subitizing* refers to a very rapid ability to enumerate accurately sets of generally fewer than four or five items on the basis of a process of visual pattern recognition. It is believed to result in discrete mental representations of numerical patterns, which differ from the more continuous representations of larger numerical values (Feigenson, Dehaene, & Spelke, 2004; Kaufman, Lord, Reese, & Volkmann, 1949; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). If mental representations of numbers less than or equal to five were discrete rather than continuous, response latencies would be brief and similar, and responding would show little to no variability or errors.

Human nonverbal absolute discriminations of produced responses or analogue and symbolic stimuli tend to resemble those of nonhumans; estimates increase linearly with number, with scalar variability and constant CVs. Accuracy increases with age and decreases with numerical magnitude. Evidence for a subitizing mechanism has been mixed, with no supporting evidence found with stimuli not presented in stimultaneous visual arrays (Cordes et al., 2001; Huntley-Fenner, 2001; Whalen, Gallistel, & Gelman, 1999).

Multiple numerical systems

Some research suggests that humans possess multiple numerical representations, which change over time and can be used selectively to optimize performance in various numerical tasks (Dehaene, Izard, Spelke, & Pica, 2008; Feigenson et al., 2004; Siegler & Booth, 2004; Siegler & Opfer, 2003). One is a nonverbal approximate estimation system that is used for representing large or unfamiliar numbers, the same as or similar to that used by nonhuman

animals, and produces scalar response variability, consistent with Weber's law. The other permits more accurate numerical discrimination, through either a perceptual subitizing-based system for small values or a different system that develops with age and numerical experience, in which both large and small numbers are represented along a linear scale with constant generalization between values. This more precise linear representation represents and discriminates between all numerical values equally, allowing for the exact discrimination of small and large numbers and the use of arithmetic operations. Consistent with this, studies have shown that children produce numerical estimates that are increasingly more linear than logarithmic and, subsequently, more accurate with age (Siegler & Booth, 2004; Siegler & Opfer, 2003).

Present experiment

Recent research has found that the response characteristics of pigeons in absolute and relative numerosity discrimination procedures can resemble those of human verbal numerical discriminations. Tan et al. (2007) trained pigeons in a numerical reproduction task. Following two, four, or six response-dependent keylight flashes in a sample phase and a brief retention interval, subjects were required to make a number of keypecks equal to the number of flashes, followed by an additional response on another key, to obtain food. Randomizing the interflash intervals and the total duration of the sample phase controlled the influence of temporal cues.

Tan et al. (2007) found that response number increased linearly with flash number and showed positive transfer to novel numbers from 1 to 7. Of particular interest, response variability was nonscalar and resembled binomial variability, where relative response variability decreased, instead of remaining constant, as number increased. This finding was similar to that obtained by Cordes et al. (2001) in a verbal counting condition and was replicated in a later experiment that examined acquisition in the same procedure (Tan, 2009). Performance in a bisection task adapted from the reproduction procedure also showed characteristics of a nonscalar numerical representation; bisection points were located at the arithmetic, not the geometric, mean (Tan & Grace, 2010). Arithmetic mean bisection and binomial variability are hallmarks of human verbal numerical discriminations and suggest a linear representation with constant generalization (Cordes et al., 2001; Gibbon, 1981). These anomalous findings provided the basis and impetus for the present research.

We studied the performance of humans in nonverbal relative and absolute numerical discrimination tasks. A major question was whether responding in the nonverbal discriminations would show bisection at the arithmetic mean and decreasing relative response variability, similar to pigeons' responding in analogous tasks (Tan & Grace, 2010; Tan et al., 2007).

We used the consecutive numbers from 1 to 20 as stimuli in this experiment; the majority of prior research with both humans and nonhumans has used nonconsecutive numbers that are generally all less than or greater than 10. A dense range of values that encompasses 10 would permit the detection of any magnitude effects on responding (e.g., discrimination of small vs. large numbers) and would identify a point where response variability changes from nonscalar to scalar. If a process analogous to subitizing is active for small numbers, variability should change at approximately 4 or 5 (e.g., Feigenson et al., 2004; Revkin et al., 2008).

Method

Participants

Participants were 24 males and females of varying ethnicity and backgrounds. They were mostly university students, and ages ranged from 20 to 47 years (mean age = 25.97). All had normal color vision and received a $20 petrol voucher for participation.

Apparatus

Experiments were conducted on a Compaq Evo PC, using E-Prime software. Responses were recorded on the keyboard. Stimuli were presented on a 15-in. LCD monitor set back approximately 22 cm from where participants were seated.

Stimuli

Stimuli were red and black images of 46 identifiable objects—for example, anchor, bottle, door, feather, saw, teapot, vase. Stimulus images were sourced from the International Picture Naming Project (http://crl.ucsd.edu/~aszekely/ipnp/1stimuli.html) and the Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University (http://www.tarrlab.org; courtesy of Michael J. Tarr). The images varied from 8 cm to 10 cm in size. Outlines and shading of the target stimuli were colored red, using Microsoft Paint or Microsoft Picture Editor. All images were used as both target and distractor stimuli.

Procedure

Participants read written instructions and procedural details before starting the experiment. The experimenter reiterated

the instructions and answered any additional questions. Participants were asked to name each picture aloud as it was presented and to keep track of the number of red pictures presented, without counting them. The numerical values or range of numbers that would be presented were not provided. The experimenter remained in the room and monitored participants' compliance with the instructions. Any failure to adhere to the instructions was corrected with a verbal reminder of the task requirements.

Each trial consisted of a sample phase and a response phase. Sample phases were initiated by a space bar press and consisted of a sequence of 40 objects colored red or black, each with a 500 ms onset and offset delay. The sample number was the number of red objects distributed randomly within the sequence. Although each image was presented regularly in time, the number of black objects between each red object (analogous to the interflash-interval in Tan et al., 2007) was randomly determined, such that the intervals between each red object varied irregularly. At the end of every sample phase, participants were presented with on-screen instructions of the requirements for the subsequent response phase.

There were three different response conditions, presented in separate blocks and counterbalanced order:

1. *Bisection*. Participants pressed the "m" key on the keyboard if the target number was a large number or "c" if it was small. On the basis of pilot testing, participants were told the anchor values at the beginning of the experiment—20 and 1 were large and small, respectively—to discourage the exclusive choice of a single response.
2. *Reproduction*. Participants reproduced the target number with "g" keypresses. Participants were asked to respond as quickly as possible, saying "gee" out loud with every response to prevent the active counting of responses.
3. *Report*. Participants typed in the target number using the number keys.

After each response, participants started the next trial by pressing the space bar. Each response condition consisted of 20 trials, 1 for each sample number from 1 to 20 inclusive, presented randomly without replacement. No feedback was provided, and participants completed the experiment at their own pace. Data from 3 participants were excluded due to their failure to comply with instructions. This appeared to be limited to particular response types, and so their data were excluded only from the affected conditions: One from the discrimination condition, 1 from the reproduction condition, and 1 from both the reproduction and report conditions.

Results

Our nonverbal manipulation seemed to be effective in preventing covert and overt counting strategies. Upon debriefing, participants indicated that the task of naming the pictures was sufficiently demanding in itself, and they often had difficulty merely keeping up with the rate of stimulus presentation.

Bisection

In the bisection condition, the proportion of *large* responses increased as sample number increased. Because participants experienced only a single trial of each numerical value, the data were averaged. A plot of response data can be seen in Fig. 1. To test for bisection point location, a two-parameter logistic function was fitted to the average data, using a least-squares criterion:

$$P(large) = \frac{1}{1 + e^{-A(n-B)}}, \tag{1}$$

where n is the sample number, and A and B are fitted parameters corresponding to the slope and bisection point of the function, respectively.

Equation 1 fit the data well, accounting for 93% of the total variance. The predicted bisection point was 11.11, closer to the arithmetic mean (10.50) than to the geometric mean of 1 and 20 (4.47) and the geometric mean calculated with all values between 1 and 20, inclusive (8.30). The slope (A) or sensitivity of the function was 0.34. A Weber fraction for the group function was calculated by dividing the difference limen (the difference of the stimulus values eliciting 75% and 25% of the *large* responses) by the bisection point and equaled 0.575.
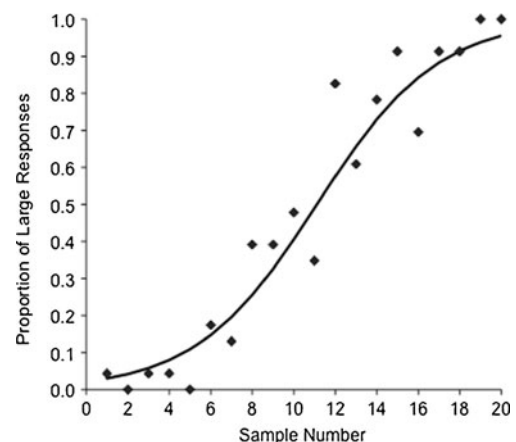

Fig. 1 Average proportion of *large* responses plotted as a function of sample number and best-fitting logistic function

We conducted Zeiler's (1985) power mean analysis to determine whether our obtained bisection point was best fit by the arithmetic or geometric mean. The power mean (*PM*) is calculated using the following equation:

$$PM = \left[\left(\sum_{j=1}^{n} X_j^a \right) / n \right]^{1/a} \qquad (2)$$

According to Eq. 1, the *PM* is the arithmetic mean of the stimulus numbers (X_j) raised to the power of *a*, raised to the power of $1/a$. The value of *a* is revealing; a value close to 1 suggests that the arithmetic mean is the best description for the data, *a* approximating 0 indicates geometric mean bisection, and *a* close to –1 suggests harmonic mean bisection. The best-fitting value of *a* for the obtained bisection point was determined by using nonlinear optimization (Microsoft Excel) and a least-squares criterion. The value of *a* for the best-fitting function was *a* = .94, which indicates arithmetic mean bisection.

Reproduction

The number of responses in the reproduction condition increased linearly with sample number. A plot of the average response number for all participants is shown in the upper left panel of Fig. 2. Response number tended to increase equivalently with sample number up to 5 or 6, but beyond

that, response estimates underestimated sample number. Consistent with this, a linear regression found a significant positive relationship between sample and response number, *B* (slope) = 0.51, β (standardized slope) = 0.69, *p* < .001.

Average performance never exceeded 85%, and for values greater than approximately 7, accuracy peaked at approximately 15%. A repeated measures ANOVA found a significant effect of number, $F(19, 361) = 7.90$, *p* < .001, but no significant effect of order (whether the reproduction condition was experienced as the first, second, or third block), $F(2, 19) = 0.10$, n.s., and no number/order interaction, $F(38, 361) = 1.25$, n.s. A trend analysis showed that the proportion of correct responses decreased linearly as sample number increased, $F(1, 19) = 64.81$, *p* < .001.

CVs were calculated by dividing the standard deviation by the mean response number for each numerical value. Mean CVs and model fits are plotted against mean response number on a log-log scale in the lower left panel of Fig. 2, in the same manner as in Cordes et al. (2001). The data show two distinct patterns; CVs decrease for the first 7 or 8 values and then increase for larger values.

Polynomial regressions were conducted to assess the fit of linear and quadratic models to the log CVs and log average response number data. A linear model did not provide a good account of the data, $F(1, 18) = 0.92$, n.s., whereas a quadratic model did, $F(2, 17) = 15.07$, *p* < .001. To identify the inflection point where the slopes changed

Fig. 2 Upper left and right panels: Average response number plotted as a function of sample number in the reproduction and report conditions, respectively. Error bars represent +1 *SD*, and the dotted line plots the line of equality. Lower left and right panels: Log coefficients of variation plotted as a function of log average response number for the reproduction and report conditions, respectively. Solid lines show the fit of the bilinear function to the data
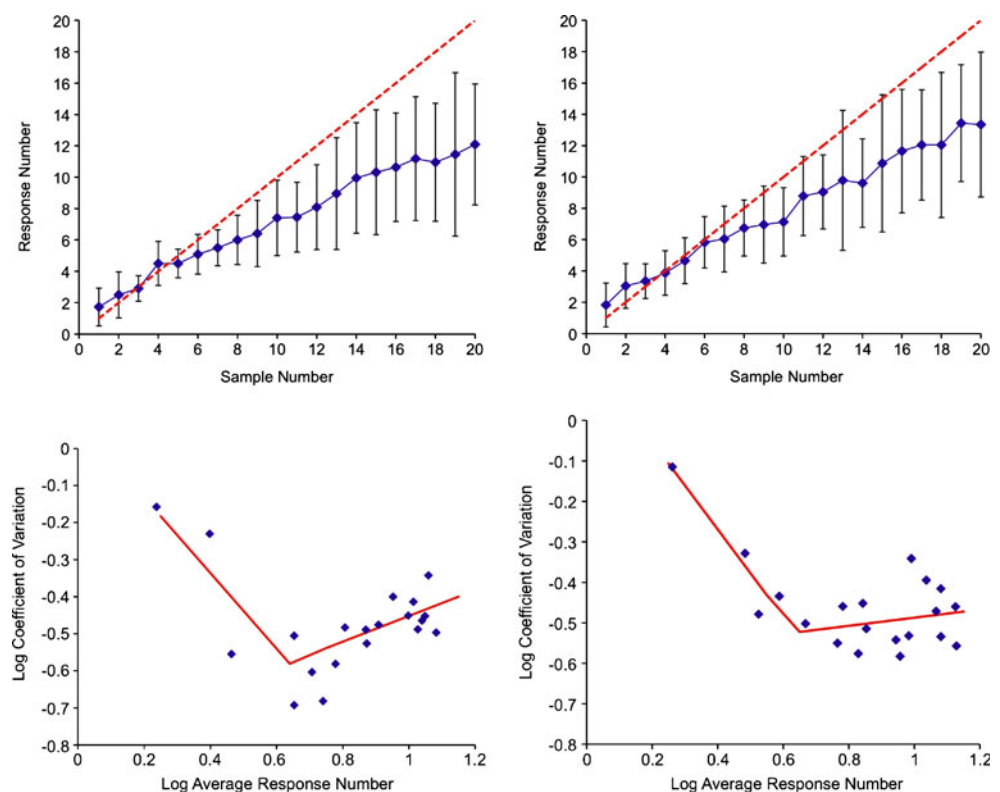
from negative to positive, a "broken stick" bilinear model was fit to the data, and its performance was compared with that of a linear model.

Slopes, intercepts, and the variance explained when predicting log CVs from response number were calculated separately for both linear and bilinear models, using nonlinear optimization (Microsoft Excel) and a least-squares error criterion. The linear model performed poorly, accounting for 5% of the variance with a slope and intercept of –0.11 and –0.38, respectively. In contrast, the bilinear model performed well, accounting for 74% of the variance. Obtained slopes for the bilinear model were –1.02 and 0.35, with intercepts of –0.07, and –0.80, for values less than and greater than the inflection point, respectively. These results were obtained with an inflection point located at 0.76 log average response number, corresponding to a flash number of approximately 7. The increase of 69.5% of the variance accounted for by the bilinear model over the linear model was significant, $F(1, 17) = 46.08$, $p < .001$.

To examine whether the obtained slopes in the bilinear model were significantly different from 0, linear regressions were used to calculate the slopes for the first seven log CV values (equivalent to flash numbers 1–7) and for the remaining values. The slope of the first seven values was significant and negative, $\beta = –0.90$, $R^2 = 80\%$, $p < .01$, whereas the slope for values greater than 7 was significantly positive, $\beta = 0.59$, $R^2 = 35\%$, $p < .05$.

Report

Response number in the report condition increased with sample number. Plots of average reported number are shown in the upper right panel of Fig. 2. Response number was approximately equal to sample number for values up to approximately 6, while for values larger than 6, sample number was consistently underestimated. A repeated measures ANOVA found a significant effect of number on responding, $F(19, 380) = 48.60$, $p < .001$, no significant effect of order, $F(2, 20) = 0.19$, n.s., and no significant interaction, $F(38, 380) = 0.75$, n.s. The slope of the function when response number was plotted against sample number was larger than in the reproduction condition, $\beta = 0.76$, $p < .001$.

Interestingly, participants showed an overall tendency to report larger numbers than in the reproduction condition (overall $M = 8.00$, $SD = 3.58$ for report, and overall $M = 7.38$, $SD = 3.23$ for reproduction); a two-way repeated measures ANOVA, with condition and number as factors, showed that this difference in response number was significant, $F(1, 21) = 13.24$, $p < .005$. No significant interaction between number and response type was found, $F(19, 399) = 1.27$, n.s.

As is suggested by Fig. 2, proportion of correct trials decreased as flash number increased. A repeated measures ANOVA revealed a significant effect of number, $F(19, 380) = 3.78$, $p < .001$, and a significant linear trend, $F(1, 20) = 41.97$, $p < .01$. There was no significant difference in overall proportion correct in the reproduction (overall $M = 0.24$, $SD = 0.21$) and report (overall $M = 0.25$, $SD = 0.15$) conditions; across all sample numbers, participants were equally accurate, regardless of response type, $t(19) = 0.36$, n.s.

Coefficients of variation for the report condition are plotted in the lower right panel of Fig. 2. CVs decreased for the first six values but did not appear to vary systematically with number for larger values. Data were analyzed in the same way as the reproduction condition. Polynomial regressions between log CV and log average response number showed significant linear, $F(1, 18) = 7.65$, $p < .05$, as well as quadratic, $F(2, 17) = 14.81$, $p < .01$, components. The slope and intercept of the best-fitting linear function were –0.24 and –0.25, respectively. For the bilinear function, the best-fitting slopes were –0.96 and 0.01, and the intercepts were 0.13 and –0.50 for values less than and greater than an inflection point of 0.52, respectively, which corresponds to a flash number of approximately 3. Note that assuming a greater inflection point within a range of 0.6–0.65, or a flash number of approximately 4 or 5, obtained similar results; $R^2 = 66\%$, slopes of –1.09 and 0.10, and intercepts of 0.17 and –0.59 for values less and greater than the inflection point, respectively. When a broken-stick bilinear model was fitted to the data using an inflection point of 0.6, the bilinear model accounted for an additional 37% of variance beyond the linear model, which was significant, $F(1, 17) = 19.07$, $p < .001$.

When an inflection point corresponding to a flash number of 5 was used, regressions of log CV on log average response number revealed a significant negative slope of $\beta = –0.95$, $p < .05$, for flash numbers 1–5 but not for values greater than 5, $\beta = 0.21$, n.s. Thus, although CVs in the report and reproduction conditions showed a similar overall pattern that was best described by a bilinear function with decreasing CVs for values less than the inflection point, some features differed. The inflection point was lower in the report than in the reproduction condition, and log CVs for average response numbers greater than the inflection point were approximately constant in the report condition but increased in the reproduction condition.

Discussion

The results from this experiment showed that humans are able to discriminate numbers in nonverbal bisection, reproduction, and report tasks. In all conditions, response

estimates increased with number and responding in the reproduction and report conditions exhibited both scalar and nonscalar variability.

One potential criticism of the present procedure is the inherent correlation between number, duration, and rate of stimulus presentation. Given the constant total duration of the whole stimulus sequence and the constant onset/offset times of the sample stimuli (red objects), the total duration and overall presentation rate of red objects in each sequence were perfectly correlated with sample number. However, for several reasons, it is unlikely that participants were responding on the basis of these temporal characteristics. First, responding based on the total duration of the sample stimuli would require the summation of the onset and offset times of each individual red object presented in an irregular sequence. Although overall rate covaried with number, the random presentation of the red objects within black ones randomized the individual intervals between any two red objects. This would have made the estimation of overall rate or duration of sample presentation much more difficult than if the red objects had occurred consecutively. Second, participants were explicitly instructed to focus on number in the instructions, and given the combination of the weaker salience of duration than number in the present procedure, the interference of number on temporal discriminations (Droit-Volet et al., 2003), and the distractor task, it seems unlikely that participants would have based responses on an extraneous, more difficult dimension. Finally, reports from participants upon debriefing suggested that they were not timing responses, and responding did not exhibit the typical characteristics of temporal discrimination responding, geometric mean bisection, and scalar variability. Thus, it seems reasonable to conclude that participants were discriminating number, not time.

It is possible that participants may have furtively used physical prompts to assist numerical discrimination, which could have influenced performance, particularly with smaller numbers. To prevent or, at least, reduce the likelihood of these behaviors, counting was verbally discouraged in the instructions, and the experimenter remained in the room and observed participants. The experimenter did not notice the use of subtle physical cues by any of the participants, and the distractor task appeared to be sufficiently demanding to prevent this. However, one must acknowledge the possibility that the use of surreptitious cues may have influenced performance. This could be minimized in future research by the use of a distractor task that requires both physical and verbal responses to stimuli.

In the bisection task, *large* responses increased with number, and the average bisection point was located closer to the arithmetic than to the geometric mean, similar to Droit-Volet et al. (2003) and Tan and Grace (2010), although inconsistent with other numerical bisection studies

(e.g., Beran et al., 2008; Jordan & Brannon, 2006a; Roitman et al., 2007). This suggests that participants used a linear number scale with constant variability, rather than one with scalar characteristics that conforms to Weber's law (Roitman et al., 2007). The group Weber fraction was high, suggesting low discriminability, and similar to that obtained with children and monkeys by Beran et al. (2008), although higher than those in other, comparable human studies (Droit-Volet et al., 2003; Jordan & Brannon, 2006a; Roitman et al., 2007).

Both Droit-Volet et al. (2003) and the present experiment required participants to discriminate the number of sequential stimuli and obtained bisection at the arithmetic mean. Jordan and Brannon (2006a) suggested that sequential stimuli might promote the development of a more serial, linear-based representation of number and responding that becomes relatively more accurate as number increases, instead of scalar variability. Consistent with this notion is the finding of scalar variability in similar reproduction and response tasks with nonsequential stimuli, such as in Whalen et al. (1999), who used Arabic numerals as indicators of response requirement, and in other studies that have found scalar variability using simultaneously presented stimuli (e.g., Beran & Rumbaugh, 2001; Emmerton & Renner, 2006; Feigenson et al., 2004). However, scalar responding has also been found in several studies using sequentially presented stimuli (e.g., Allan & Gibbon, 1991; Boisvert, Abroms, & Roberts, 2003; Breukelaar & Dalrymple-Alford, 1998; Fetterman, 1993; Meck & Church, 1983; Platt & Johnson, 1971; Roberts, 2005), and scalar response variability was found for larger numbers in the present experiment, so mode of stimulus presentation cannot be the sole explanation for these findings.

Responding in the reproduction and report conditions shared key features: (1) Average response number increased with sample number, although it was increasingly underestimated, and (2) Coefficients of variation decreased for values less than approximately 6 and then increased or remained constant for larger values. This pattern was more marked in the reproduction than in the report task, suggesting that variability was due largely to errors in the discrimination of response, rather than to stimulus number. It is unclear what is responsible for this effect. It may be a result of the compression of the numerical scale, or memorial decay, which is likely when verbal intermediaries between stimuli and response are unavailable.

Previous research using a Mechner (1958) type response production procedure with pigeons (Machado & Rodrigues, 2007) also obtained decreasing CVs, although for a larger range, up to approximately 10. Consequently, it would be reasonable to hypothesize that this might be a result of the response phase structure. However, we found this result in both the reproduction and report conditions, suggesting that

the changes in relative response variability were not just an artifact of the reproduction task but may reflect differences in processing over different numerical ranges.

Our finding of heterogeneous response variability has implications for the structure of the underlying numerical representation developed in this procedure. The finding of binomial variability suggests that for values 1–5, at least, responding is based on a linear scale with constant generalization between values, rather than on a logarithmic scale with constant generalization or a linear scale with scalar generalization, both of which would predict scalar variability. However, the scalar variability observed for values larger than about 6 supports the latter two scales.

The response variability patterns that we observed suggest that participants may use multiple representations of number, consistent with research that suggests that such representations may change and develop with age and experience (Revkin et al., 2008; Siegler & Booth, 2004; Siegler & Opfer, 2003). Siegler and Booth proposed the use of a linear scale primarily for the discrimination of familiar small numbers and a logarithmic scale with larger numbers, which are represented inexactly. It is possible that the increased difficulty of nonverbally discriminating and reproducing larger numbers elicited the use of a logarithmic numerical scale.

The failure of previous studies to obtain evidence for different scales within the same task might have been due to the use of numerical values that were located in the range of only one of these scales, and not across both. For example, Whalen et al. (1999) used larger values between 7 and 25, while Revkin et al. (2008), who specifically examined differences in the discrimination of smaller versus larger numbers, used two separate ranges: all values between 1 and 8 and the decade numerosities between 10 and 80. However, Cordes et al. (2001) tested consecutive values less than 6, as well as nonconsecutive values that spanned the subitizing–nonsubitizing threshold up to 32, and notably did not find any difference in responding CVs across all the numerical values. At this point, we are unable

to explain the discrepancy in CV patterns across studies. When numerical values are spaced further apart, the resolution of variability analyses is reduced; future research needs to examine wider ranges of numbers with smaller spacing in order to fully investigate changes in variability and numerical processes as a function of number.

Cross-species comparisons

We compared the participants' performance with that of pigeons in previous numerical discrimination experiments (Tan, 2009; Tan & Grace, 2010; Tan et al., 2007), using response data for the numerical values common to the human and pigeon experiments (1–7). There were strong parallels in the performance of pigeons and human participants in these tasks, hinting at potentially similar underlying processes.

Logistic functions provided good fits of both human and pigeon data in the discrimination task. Although different number ranges were used in these experiments, the general forms of the psychometric plots were similar. Figure 3 shows a representative psychometric plot for the pigeons from one condition in a discrimination procedure (left panel; adapted from Tan & Grace, 2010) and psychometric plots for the human participants (right panel) in the nonverbal discrimination condition. Notably, bisection points were located closer to the arithmetic mean than to the geometric mean in all conditions, regardless of species and whether discrimination was verbal or nonverbal. The Weber fraction for humans was larger than that obtained for pigeons (see Tan & Grace, 2010), suggesting that pigeons were more sensitive to number, which is to be expected given their greater exposure to the bisection procedure.

The average response number of pigeons and humans in the reproduction and report conditions are plotted in the left panel of Fig. 4. Human responding in both the reproduction and report conditions resembled pigeons' responding in reproduction tasks over the same numerical range, although pigeons showed greater overestimation of smaller numbers

Fig. 3 Average proportion of *large* responses plotted as a function of sample number, on a relative scale for pigeons (left panel; data from Tan & Grace, 2010) and for humans in the nonverbal group (right panel)
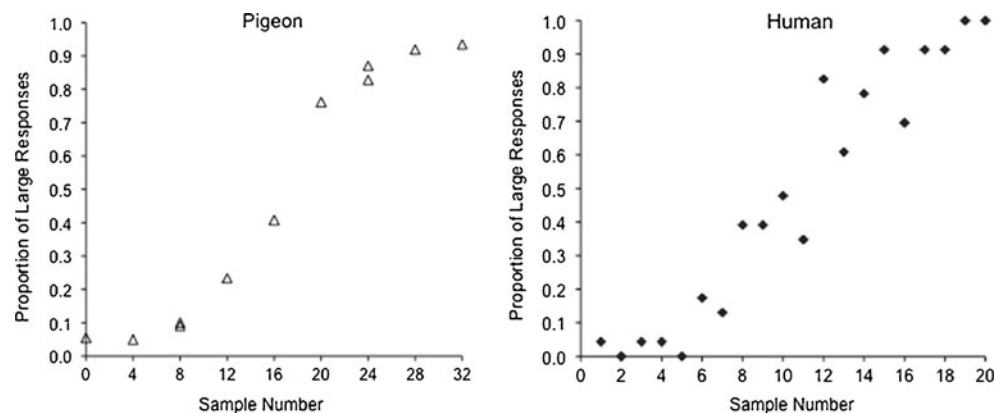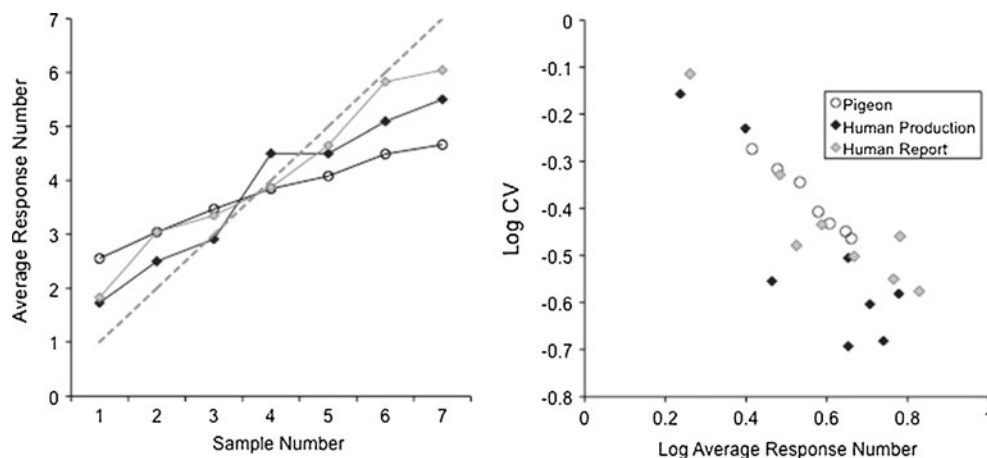
Fig. 4 Left panel: Average response number obtained for pigeons (open circles) and for humans in the nonverbal reproduction (light gray diamonds) and nonverbal report (black diamonds) conditions. Right panel: Log CVs plotted as a function of log response number for the same groups

and greater underestimation of larger numbers than humans (cf. Lejeune & Wearden, 2009). This may be due to the greater experience of human participants with number and the explicit instructions to respond on the basis of number, which has been shown to increase numerical sensitivity (Droit-Volet et al., 2003; Roitman et al., 2007). It is possible that if participants had learned task requirements through trial and error, like pigeons in the reproduction procedure, responding would show a similar reduction in numerical sensitivity and greater scale compression.

Obtained CVs for pigeons and humans were similar. Log plots of CVs as a function of average response number are shown in the right panel of Fig. 4. To compare CVs across species, regression analyses were used to calculate slopes and intercepts. The slope for the average log CVs for pigeons was significant and negative, $\beta = -0.94$, $p < .01$. The slopes for the corresponding first seven log CV values from the present experiment were also significant and negative, $\beta = -0.90$, $p < .01$, and $\beta = -0.89$, $p < .01$, for the reproduction and report data, respectively.

The log CVs obtained for pigeons were regressed onto both the human report and reproduction log CVs, and significant positive relationships were found both between pigeon and human report CVs, $\beta = 0.78$, $R^2 = 61\%$, and between pigeon and human reproduction CVs, $\beta = 0.89$, $R^2 = 78\%$. These findings suggest that although pigeons showed greater compression in the response number scale than human participants did, relative variability patterns for both species were similar.

The findings of arithmetic mean bisection and strongly decreasing relative variability for values 1–7 for both pigeons and humans in the reproduction and report tasks suggest common numerical processes and a linear numerical representation with constant generalization between values. This is noteworthy because it differs from the majority of previous studies on nonverbal numerical discriminations with both nonhumans (e.g.,

Emmerton & Renner, 2006; Fetterman, 1993; Meck & Church, 1983) and humans (e.g., Cordes et al., 2001; Whalen et al., 1999), which have found bisection at the geometric mean and scalar variability. The decreasing response variability does not appear to be an artifact of the reproduction task, since similar patterns are obtained when participants are required only to report the sample number.

These results suggest that nonverbal discrimination of number in humans and that in pigeons differ quantitatively, but not qualitatively, in the present study and previous research (Tan & Grace, 2010; Tan et al., 2007), implying a continuity of process between species with very different evolutionary histories. Humans are more accurate, but the qualitative characteristics of responding—a linear relation with sample number and decreasing relative variability—are similar.

Conclusions

Human nonverbal discrimination of both absolute and relative number was similar to that of pigeons in analogous tasks. Analyses of response variability suggest that there may be two separate processes or representations operating across different ranges, producing binomial variability for values less than approximately 6 and scalar variability for values greater than 6. This is consistent with previous research (e.g., Siegler & Booth, 2004), which has found that humans were able to use both logarithmic and linear number scales within the same task selectively, depending on the context. This suggests that the difference between verbal and nonverbal numerical processing might not be discrete. The nature of the representation and discrimination of number may be more fluid, and procedural limitations such as restricted numerical ranges and restricted response measures may have prevented previous studies from

observing both processes. Further research that draws closer parallels between human and nonhuman, or verbal and nonverbal, numerical tasks may be critical to understanding the many facets of numerical ability.

References

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation, 22,* 39–58.

Beran, M. J., Johnson-Pynn, J. S., & Ready, C. (2008). Quantity representations in children and rhesus monkeys: Linear versus logarithmic scales. *Journal of Experimental Child Psychology, 100,* 225–233.

Beran, M. J., & Rumbaugh, D. M. (2001). "Constructive" enumeration by chimpanzees (*Pan troglodytes*) on a computerized task. *Animal Cognition, 4,* 81–89.

Boisvert, M. W., Abroms, B. D., & Roberts, W. A. (2003). Human nonverbal counting estimated by response production and verbal report. *Psychonomic Bulletin & Review, 10,* 683–690.

Breukelaar, J. W. C., & Dalrymple-Alford, J. C. (1998). Timing ability and numerical competence in rats. *Journal of Experimental Psychology. Animal Behavior Processes, 24,* 84–97.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review, 8,* 698–707.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the numbers scale in Western and Amazonian indigene cultures. *Science, 320,* 1217–1220.

Droit-Volet, S., Clement, A., & Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: A developmental approach. *Journal of Experimental Child Psychology, 84,* 63–76.

Emmerton, J., & Renner, J. C. (2006). Scalar effects in the discrimination of relative numerosity in pigeons. *Learning & Behavior, 34,* 176–192.

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

Fetterman, J. G. (1993). Numerosity discrimination: Both time and number matter. *Journal of Experimental Psychology. Animal Behavior Processes, 19,* 149–161.

Gibbon, J. (1981). On the form and location of the psychometric bisection function for time. *Journal of Mathematical Psychology, 24,* 58–87.

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5–7 year olds. *Cognition, 78,* B27–B40.

Jordan, K. E., & Brannon, E. M. (2006a). A common representational system governed by Weber's law: Nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. *Journal of Experimental Child Psychology, 95,* 215–229.

Jordan, K. E., & Brannon, E. M. (2006b). Weber's law influences the numerical representations in rhesus macaques (*Macaca mulatta*). *Animal Cognition, 9,* 159–172.

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

Lejeune, H., & Wearden, J. H. (2009). Vierordt's The experimental study of the time sense (1868) and its legacy. *European Journal of Cognitive Psychology, 21,* 941–960.

Machado, A., & Rodrigues, P. (2007). The differentiation of response numerosities in the pigeon. *Journal of the Experimental Analysis of Behavior, 88,* 153–178.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *Journal of the Experimental Analysis of Behavior, 1,* 109–121.

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

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogenous behavior chain: Effect of error contingencies. *Learning and Motivation, 2,* 386–414.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19,* 607–614.

Roberts, W. A. (2005). How do pigeons represent numbers? Studies of number scale bisection. *Behavioural Processes, 69,* 33–43.

Roberts, W. A., & Mitchell, S. (1994). Can a pigeon simultaneously process temporal and numerical information? *Journal of Experimental Psychology. Animal Behavior Processes, 20,* 66–78.

Roitman, J. D., Brannon, E. M., Andrews, J. R., & Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychologica, 124,* 296–318.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75,* 428–444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14,* 237–243.

Tan, L. (2009). *The discrimination and representation of relative and absolute number in pigeons and humans.* Unpublished doctoral thesis, University of Canterbury.

Tan, L., & Grace, R. C. (2010). Discrimination and representation of relative numerosity in a bisection task by pigeons. *Learning & Behavior, 38,* 408–417.

Tan, L., Grace, R. C., Holland, S., & McLean, A. P. (2007). Numerical reproduction in pigeons. *Journal of Experimental Psychology. Animal Behavior Processes, 33,* 409–427.

Wearden, J. H., & Ferrara, A. (1995). Stimulus spacing effects in temporal bisection by humans. *Quarterly Journal of Experimental Psychology B, 48B*(4), 289–310.

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

Zeiler, M. D. (1985). Pure timing in temporal differentiation. *Journal of the Experiment Analysis of Behavior, 43,* 183–193.