Ordinality and Inferential Abilities of a Grey Parrot (Psittacus erithacus)

Irene M. Pepperberg
Brandeis University

Ordinality and Inferential Abilities of a Grey Parrot (*Psittacus erithacus*)

Irene M. Pepperberg[1,2,3]

[1] *Brandeis University*
[2] *Radcliffe Institute for Advanced Study*
[3] *Harvard University*

KEYWORDS

grey parrot, Arabic numerals, ordinality, equivalence

ABSTRACT

A grey parrot (Psittacus erithacus), able to label the color of the bigger or smaller object in a pair (I. M. Pepperberg & M. V. Brezinsky, 1991), to vocally quantify ≤6 item sets (including heterogeneous subsets; I. M. Pepperberg, 1994), and separately trained to identify Arabic numerals 1–6 with the same vocal English labels but not to associate Arabic numbers with their relevant physical quantities, was shown pairs of Arabic numbers or an Arabic numeral and a set of objects and was asked for the color of the bigger or smaller one. The parrot's success showed he (a) understood number symbols as abstract representations of real-world collections, (b) inferred the relationship between the Arabic number and the quantity via stimulus equivalence, and (c) understood the ordinal relationship of his numbers.

The comparison of nonhumans' number concepts with those of human children continues to be a vibrant research topic. According to some researchers, animals' number sense is a neurally based, evolutionarily inherent skill (e.g., Dehaene, 1997; Gallistel & Gelman, 1992) from which human competence derives. Recent studies on possible parallels in brain areas underlying numerical judgments in humans and monkeys (review in Göbel & Rushworth, 2004; see also Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, Piazza, Pinel, & Cohen, 2003; Nieder & Miller, 2003) support such an evolutionary continuum in numerical abilities.[1] Nevertheless, more advanced numerical abilities— exact counting of quantities ≥4 and arithmetic operations— have been considered to be uniquely human and based on language skills (Spelke & Tsivkin, 2001). Although primates trained with human-based symbolic systems involving Arabic numerals (chimpanzees, *Pan troglodytes*; Boysen, 1993; Boysen & Berntson, 1989, 1990; Boysen & Hallberg, 2000; Matsuzawa, 1985) and nonprimates' referential use of vocal English number words (Grey parrots, *Psittacus erithacus*; Pepperberg, 1987, 1994, 2006; Pepperberg & Gordon, 2005) might seem to provide intermediary links between animals that lack such training and children, questions still remain as to the extent of animal number competence compared to that of humans (e.g., Mix, Huttenlocher, & Levine, 2002). In particular, full *symbolic* understanding of ordinality seems difficult for nonhumans to master.

Ordinality is important because it is intrinsic to verbal counting (Davis & Pérusse, 1988; Fuson, 1988; Gelman & Gallistel, 1986). To count, an organism must produce a standard sequence of number tags and know the relationships among and between these tags; for example, that "two" (be it any vocal or physical symbol) not only comes before "four" in the verbal sequence but also represents a quantity less than "four." Children seem to integrate their acquisition of ordinal– cardinal abilities in a stepwise manner. They first learn cardinality for very small numbers (<4) and a general sense of "more versus less," next acquire a meaningless, rote ordinal number series, then associate their knowledge of quantity in the small sets with this number sequence to form 1:1 correspondences that can be extended to larger amounts for both cardinal and ordinal accuracy (e.g., Brannon, 2002; Carey, 2004; Xu & Spelke, 2000; note Liddle & Wilkinson, 1987, review in Mix et al., 2002). It is interesting to note that children may learn associative rules that allow them to act as though they have full understanding of the cardinality of number words or numerals before they have such competence (i.e., respond correctly to "How many?" but fail on "Give me X"; see Sophian, 1995) but cannot do so with ordinality (Bruce & Threlfall, 2004; Teubal & Guberman, 2002).

Ordinality studies are also complicated by the fact that most experiments on young children and animals involve ordering collections of objects (e.g., Brannon, 2002; Brannon & Terrace, 1998, 2000; Emmerton, Lohmann, & Niemann, 1997; Judge, Evans, & Vyas, 2005; Thomas, Fowlkes, & Vickery, 1980) and do not involve *symbolic* ordering of Arabic numbers based on their referential value (e.g., Brannon, Wusthoff, Gallistel, & Gibbon, 2001). Rhesus monkeys (*Macaca mulatta*; Washburn & Rumbaugh, 1991) and squirrel monkeys (*Saimiri sciureus*; Otlhof, Iden, & Roberts, 1997) learned to order Arabic numbers based on quantities to which they referred and to infer relations among novel pairings, but symbolic ordering separate from numerical food rewards did not emerge spontaneously. Replication of these studies with pigeons (*Columba livia*; Olthof & Roberts, 2000) showed the birds responded to mass of food rather than number, and such controls were not always used in the monkey studies. Boysen and colleagues (Boysen, Berntson, Shreyer, & Quigley, 1993) trained chimpanzees to pick the bigger Arabic numeral of various adjacent pairs based, as were the monkey studies, on the quantities to which they referred, then showed that their subjects used transitive inference to choose the correct response for nonadjacent pairs; the need for significant training on adjacent pairs, however, showed that ordinal relationships did not emerge from the apes' previous referential use of Arabic numerals. Similarly, Matsuzawa and colleagues (Kawai & Matsuzawa, 2000; Matsuzawa, Itakura, & Tomonaga, 1991; Tomonaga, Matsuzawa, & Itakura, 1993) had to train their chimpanzee, Ai, in ordinal use of Arabic numerals she could use cardinally. The question thus arose as to whether ordinality would emerge or have to be taught to a Grey parrot, Alex, that had learned to quantify sets of objects with vocal English labels, including heterogeneous collections (e.g., the number of red balls in a collection of red and blue balls and blocks; Pepperberg, 1994). Note, he had not learned number labels in ordinal order, having first been taught "three" and "four," then "two" and "five," next "six," and lastly "one" (Pepperberg, 1987, 1994), and he might have, like young children and apes, associated number labels and quantity without understanding the symbolic ordinality of his number labels.

A task involving equivalence relations can test this ability. After being trained to identify Arabic numerals in the absence of the physical quantities to which they refer, could Alex use the commonality of English-based labels to correlate (form equivalence relations between) quantity and Arabic numerals (see Figure 1), so as to identify the *color* of one of a pair of Arabic numbers (e.g., a blue 3, a red 4, set next to each other on a tray) that is numerically (not physically) bigger or smaller? He could already answer "What color/matter [is] bigger/smaller?" for object pairs and to respond "none" if nothing was bigger or smaller (Pepperberg & Brezinsky, 1991). To succeed on this new task, he would have to choose which Arabic numeral is bigger or smaller based on deductions and inferences: Deduce that an Arabic symbol has the same numerical value as a *vocal label*, compare *representations* of quantity for which the label stands, infer rank ordering based on these representations, then state the result *orally*. Specific stimuli within

pairs would, unlike in other nonhuman studies, not have been associated with reward of the corresponding number of items (Wynne, 1992), and by requiring color, not number, label responses and querying about both larger and smaller values, use of rote replies would be less likely for a given pair (see below); pairs could be presented at least a few times with less chance of training.[2] The task would require working and long-term memory (Geary, Hamson, & Hoard, 2000) and a form of equivalence: Only by using the commonality of English-based labels to correlate—form equivalence relations between— quantity and Arabic numerals could Alex tell which of the Arabic numerals was the bigger or smaller, that is, rank order his numerals. Note that Alex had heard younger birds' training on labeling the number of sequential sounds (training to respond to, e.g., hearing three computer-generated clicks with the vocal label "three"; Pepperberg, 2006), but these labels also were not taught in sequential order, and Alex's use of such information ("more" noise) would, moreover, involve cross-modal transfer.

Figure 1. The arrow from "'vocal number label" to "object collection" represents vocal production of number labels in the presence of a collection of objects (accuracy 83.3%; Pepperberg, 1994); the arrow from "object collection" to "vocal number label" is comprehension of vocal number labels in the presence of a collection of objects (88%; Pepperberg & Gordon, 2005); the arrows represent a symmetric relationship, and Alex can do both. The arrow from "vocal number label" to "Arabic numeral" represents vocal production of number label in the presence of an Arabic number symbol; the arrow from "Arabic numeral" to "vocal number label" represents comprehension of the vocal number label in the presence of an Arabic number symbol; the arrows represent a symmetric relationship, and Alex can do both ("83%; Pepperberg, 1997). The paths between "object collection" and "Arabic numeral" are untrained. If Alex sees an Arabic numeral, will he understand that it represents the appropriate collection of objects? The arrows between "Arabic numeral" and "object collection" represent an emergent equivalence relationship (see Devany, Hayes, & Nelson, 1986).
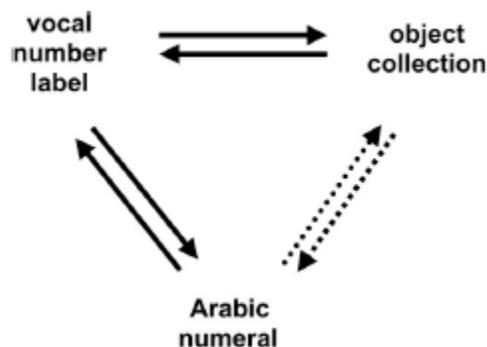
Although numbers involved in this study are relatively small (!6), the task would also show if Alex, like humans (Geary et al., 2000; Moyer, 1973), has more difficulty comparing Arabic numbers close in numerical value than those differing by larger amounts and if larger valued pairs (e.g., 5:6) present more problems than smaller ones (1:2). That is, will Alex demonstrate humanlike numerical distance effects (Buckley & Gillman, 1974; Dehaene, 1997; Moyer & Landauer, 1967)? Humans, supposedly, use internal representations (maps to exact quantities or number lines) to determine which of two numerals is bigger or smaller; they thus find the task more difficult for close numbers—even for small quantities— because representations overlap more. Most human studies compare reaction times to determine these effects, but such is not possible with Alex. His reaction time depends on his immediate interest in the task (Pepperberg, 1999); however, error rates may provide a measure of difficulty. The data could also be analyzed to see if Alex had more difficulty if numerals were not in ordinal order on the tray; some

evidence suggests that such ordering slows reaction times in humans (Dehaene, Bossini, & Giraux, 1993). Results may thus provide some means of comparing his symbolic understanding of number labels to that of humans (Whalen, Gallistel, & Gelman, 1999).

Three other types of comparisons were also of interest. First, to see if physical size might confound symbolic issues for Alex as it sometimes does for young children (e.g., Girelli, Lucangeli, & Butterworth, 2000; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002), I presented trials with identical numerals of different sizes and colors (e.g., 6:6). Children given such exemplars were told when the target of interest was size versus number, and I could not, without significant additional training (on the label "size"),[3] similarly alert Alex, but I could determine on which basis he would preferentially respond; that is, would he say "none" or the color of the physically bigger or smaller numeral? Second, trials on identical numerals of different colors but of the same size (e.g., 6:6) were included as controls for this test; that is, to see if Alex would, as expected, reply "none" to the query as to which was bigger or smaller. Third, in a small number of trials, I mixed Arabic symbols and physical items (one for each possible pairing, including identical numerical values). Numerals were placed next to random arrays of objects on a tray, counterbalanced with respect to left–right, bigger–smaller queries; Alex was again to label the color of the bigger or smaller numerical quantity. A correct response to, for example, one numeral (an Arabic 6) as bigger than five items (or an Arabic 2 as the same as two items) would show further understanding of the abstract nature of symbols and test ordinality of numerosities in a manner cleanly separating mass and number (for children's confounds of mass and number, see Mix et al., 2002; Sophian, 2000). Such trials would be another control for learning a specific response to specific numeral pairs. A few trials were added in which pairings were such that the Arabic numeral was indeed the smaller numerical quantity, to ensure that Alex did not extract a pattern inherent in the questioning; that is, assume that the Arabic number would always be the larger quantity and then respond to bigger–smaller on that basis. The data would provide additional information on how a bird balances size with the distance effect; that is, would a distance effect exist and, if so, would it be more pronounced for symbolic or physical closeness (Rubinsten et al., 2002)?

General Method

Subject and Housing

Alex, a 28-year-old male Grey parrot (*Psittacus erithacus*), had been the subject of numerous cognitive and communicative studies for 27 years, including those involving numerical competence (Pepperberg, 1987, 1994, 2006; Pepperberg & Gordon, 2005). Testing locations and living conditions when neither testing nor training was in progress are described in Pepperberg and Wilkes (2004) for the majority of testing. The first 18 trials occurred in 1997, and living conditions are described in Pepperberg (1994). Food and water were available at his vocal request at all times. In this study he used previously documented abilities (Pepperberg, 1994, 1999; Pepperberg & Gordon, 2005) to produce and comprehend vocal labels for quantities (≤6), Arabic numbers (≤6), and colors (rose [red], green, purple, blue, yellow, orange, gray) and to understand the concept of relative size and absence of size difference (Pepperberg & Brezinsky, 1991). At the time of the testing, Alex had never received explicit training (match-to-sample or model/rival; Pepperberg, 1981) on relationships between any pairs of tested items or between object collections and Arabic numerals.

Apparatus and Materials

Tests involved familiar objects. The tray forming the substrate for all trials had been used for earlier studies on label comprehension (Pepperberg, 1990, 1992), object permanence (Pepperberg & Kozak, 1986), and other number capacities (Pepperberg, 1994, 2006, Pepperberg & Gordon, 2005). I, along with other trainers, concurrently used the tray for training phoneme recognition and spatial concepts (e.g.,

Pepperberg, in press) so it would not cue a number task. Alex had previously been trained to identify the Arabic numbers (either colored plastic magnets or food-dyed wooden numerals of equivalent size) by being given a tray of all numerals (1–6) and being queried "What color X?" (X = an Arabic numeral from 1 to 6) or "What number Y?" (Y = one of seven colors). Objects used for comparison with numerals were also familiar: pompons, wooden blocks, and so forth (see the Appendix).

General Procedure: Object Presentation

In the absence of any prior training, a trial began when a human, out of Alex's line of sight, placed either two Arabic numerals or an Arabic numeral and a collection of objects onto the tray, counterbalancing the larger–smaller numeral or set on the right and left across trials. Numerals were placed about 2.5 cm apart; for trials with object sets, objects were generally less than 1 cm apart. The human brought the tray up to Alex's face and made eye contact but did not look at the tray; Alex was then asked, vocally, to respond to "What color[ed] number [is] bigger (smaller)?" If he did not respond within approximately 30 s, the question was repeated. Trials were counterbalanced as to the question and also randomized with respect to color labels and pairings of numerals.

The experiment consisted of 110 trials. Each possible nonidentical pair of same-sized numerals (e.g., 1:6; 3:4) was presented four times, in random order (60 trials) with the smaller to the left half of the time and the other way around the other half of the time and right–left order randomized; identical pairs of same-sized numerals (e.g., 6:6) were presented twice (12 trials); identical pairs of different-sized numerals (e.g., 6:6) were presented twice (12 trials, once with the smaller on the left and once on the right; numerals differed by about 5–7 mm), and 26 trials involved objects versus symbols (see the Appendix for details of object–symbol pairings). All trials were intermixed randomly. The number of trials of any one type was deliberately kept low to avoid training. Trials somewhat overlapped a study on addition (Pepperberg, 2006), in which Alex demonstrated approximately 90% accuracy in quantifying groupings of items.

General Testing Procedure

Test sessions of a single ordinality question occurred approximately once per week between June 1997 and December 1997 and an average two to five times per week during November 2004 to June 2005, with breaks for student vacations, intersessions, and other periods of limited staffing. Only 18 trials were presented in 1997; the study was paused because of lack of staffing and several moves of the entire laboratory. Test questions were presented intermittently either during free periods (when birds would request various foods or interactions) or during sessions on current (and thus unrelated) topics (e.g., using Alex to assist in training another parrot on shape labels) until all questions for the experiments were presented. Details of test procedures, including descriptions of precautions against inadvertent and expectation cuing, are in Pepperberg (1981, 1990, 1994); summaries are below. A question would be repeated in a session only if Alex's initial answer was incorrect (e.g., Pepperberg, 1981, 1987). Specifically, the number of times an array was presented to Alex depended on his accuracy. If Alex produced the appropriate label, he received praise and the objects on the tray or was allowed to request an alternative reward (as the study progressed, chewing numerals was less rewarding; he then stated "I want X"; X usually was a cork, a nut, etc.). No further presentations of the same exemplars then occurred; that is, there was only a single, *first-trial* response. If an identification was incorrect or indistinct, the examiner removed the tray of objects, turned his or her head, and emphatically said "No!" Only under this condition (i.e., an erroneous response) were test materials immediately, repeatedly presented to penalize a win–stay strategy. Presentation continued until a correct identification was made or four attempts occurred;[4] errors were recorded.

As in all studies with Alex, protocol differed from those used with other animals in two respects. First, the task capitalized on Alex's ability to work in the vocal mode; second, each trial was generally presented intermittently during training and testing of other unrelated topics also under study. Alex's responses thus had to be chosen from his entire repertoire ($90 utterances, including food and location labels) and from among numerous possible topics concerning various exemplars and queries in each session. This design not only increased the task complexity but also prevented several forms of cuing discussed below (see also Pepperberg, 1999; Premack, 1976, p. 132).

Controls Against Expectation and Experimenter-Induced Cuing

Test situations included specific precautions to avoid cuing. One control was a design such that each test session was, as noted above, presented intermittently during free periods or work on unrelated topics. As noted earlier, Alex's responses had to be chosen from his entire repertoire and from among numerous possible topics during each session; each session contained only a single number array. An examiner who, for example, poses a series of similar questions may come to expect a particular answer and unconsciously accept an indistinct (and by our criteria incorrect) response of, for example, "gree" (a mix of "green" and "three") for "green." Second, in general, a human other than the one presenting the tray (one of four possible individuals in these experiments), who did not know what was on the tray, confirmed the answer; his or her interpretation of Alex's response was thus unlikely to be influenced by an expectation of a certain color label. Only after his or her confirmation was Alex rewarded (Pepperberg, 1981). Third, because Alex had not been trained on this task, no overlap occurred between training and testing situations and, because training on number and color labels had occurred years before with trainers no longer in the lab, Alex could not have picked up on trainer-induced cues specific to a given label (Pepperberg, 1981); moreover, because a total of five different humans were involved in the testing, the presence of a particular individual could not cue a number session. Fourth, the evaluator was unlikely to be influenced by hearing the type of question posed: In a previous study, transcriptions of contextless audiotapes of Alex's responses in a session agreed with original evaluations to within 98.2% (Pepperberg, 1992).[5] About 10% of trials were videotaped to check for interobserver reliability. If interobserver reliability with a blind coder was high, I could be assured of the validity of other trials. Given that Alex was far more interested in the student performing the taping and the video camera rather than the objects to be labeled, such trials were difficult to execute and were kept to a minimum. Note that about a third of these video trials had to be discarded because Alex would not attend to the exemplars, and of the remaining trials, one had to be discarded because Alex's speech was too soft to be heard on the tapes. Thus only six trials were available for analysis, and two of the four evaluators of these trials said Alex's speech was too soft to be understood except on one or three trials, respectively. Evaluators were people who had never been in the laboratory and who had no familiarity with Alex.

Scoring

Alex's test scores were calculated two ways. Because the test procedure required that a question could be repeated (up to four times) until Alex produced the correct response, I scored both *first-trial* and *all-trial* responses. First-trial results were the percentage of correct responses on first trials. The test score result for all trials was obtained by dividing the total number of correct identifications (i.e., the predetermined number of collections) by the total number of presentations required to obtain the correct responses. All-trial scores are reported because the type of error Alex made could provide information on exactly how he was processing the information; that is, was he consistently choosing the wrong alternative or was he having difficulty identifying a color? Except as noted, statistics were performed only on first-trial data.

To learn if Alex's results were statistically significant, I used, except where noted, a binomial test with chance of 1/2; that is, a choice between one of the two possible color labels of the numerals on the tray.

A less conservative calculation would include the probability of producing any possible color label (1/7) or any of the "90 labels in his repertoire; calculations assume that Alex will always ($p = 1$) attend and respond correctly to the "What color . . . ?" question (i.e., not provide a random label that had no connection to the task at hand). Occasionally, Alex would produce a series of color labels not on the tray; such trials (approximately 5% of the total) were, as explained in Pepperberg and Gordon (2005), discarded and repeated at a later date. If Alex had two consecutive sessions of such noncompliance, I switched to testing on a different topic for four consecutive sessions.

Results

For trials involving two different Arabic numbers of the same physical size, Alex was capable of inferring ordinal ranking. Even the errors he made provide insight into his understanding of ordinality. Of particular interest were scores the first time he was presented with each individual pair, accuracy for pairings differing by only one or two units compared with those differing by larger amounts, responses to both differently sized and same-sized identical numerals, responses on symbol versus object collection trials, and interobservability scores.

Responses for Numeral–Numeral Trials

I calculated Alex's responses for all numeral pairs (60 trials) plus those for identical numerals of both different and equal sizes (24 trials) with respect to two different error criteria (see Table 1). Under the first, I counted as errors his initial correct response of the number label rather than the color label (e.g., if shown a blue 4 and a red 6, asked "What color bigger?" and he responded "six," "six" was technically correct, but not with respect to the actual query); his first trial score was 63/84, or 75% ($p < .01$, binomial test, chance of 1/2). For all trials (i.e., including corrections trials), his score was 84/108, or 77.7%. Initial errors were divided mostly equally between right and left presentations of the smaller value. Under the second criterion, I did not count the number label responses as errors; his first trial score was then 74/84, or 88.1% ($p < .001$, binomial test, chance of 1/2). For all trials, his score was 84/95, or 88.4%. Again, his initial errors were essentially divided fairly equally between right and left presentations of the smaller values.

Initial Trials Versus Final Trials

Alex's score for the first trial for each unique number pair, including the first set of identical numerals (same size, same value) was 15/21, or 71.4% ($p = .04$, binomial test, chance of 1/2), if I include as errors trials in which he gave the correct number rather than the color label; if I do not, his score is 19/21, or 90.5% ($p < .001$). He erred by saying "none" for a 1:2 discrimination and giving the larger rather than the requested smaller answer for a 1:5 discrimination. His score for the final trials of the same unique number pairs was 19/21, or 90.5% ($p < .001$); he did not err by giving any number labels. One error was a yellow–orange confound, the other was to state a color not on the tray. The differences in scores (even with the number label errors) are not statistically significant (Fisher's exact test, $p = .12$, one-tailed).

Distance and Size Effects

I separately analyzed Alex's response for close pairings to examine distance effects (see Table 1). If I do not count as errors instances in which he produced the correct number label rather than the color label, for numeral pairs that differed by a single unit (e.g., 3:4), Alex was correct 14/20, or 70% ($p = .06$), which just misses statistical significance. It is interesting that, of his six errors, three were to say "none," one was choosing the wrong value, and two were yellow– orange confounds (saying yellow when there was nothing yellow on the tray and the answer was orange; Alex sometimes confuses these two colors because human and parrot color perception differ somewhat;[6] see discussion in Bowmaker, Heath, Das,

& Hunt, 1994; Bowmaker, Heath, Wilkie, Das, & Hunt, 1996; Pepperberg, 1994, 2006). If I omit the color confound, $p < .01$. For numeral pairs that differed by two units (e.g., 1:3), Alex was correct 13/16, or 81.3% ($p = .01$, binomial test, chance of 1/2); if I do not count as errors his stating the correct number label rather than the color, his score is 15/16 ($p < .01$, binomial, chance of 1/2); his one error was another yellow– orange confound. For pairs that differed by four or five units (1:6, 2:6, 1:5), Alex was correct on 10/12 trials (no labeling of numeral rather than color), or 83.5% ($p = .02$, binomial test, chance of 1/2). One error was to state "none," the other was to give the color of the bigger, not the smaller, numeral. The difference between trials (including the color confounds) involving large (four or five units) and small (one unit) differences in numerals was nonsignificant (Fisher's exact test, $p = .34$, one-tailed).

Table 1. Summary of Results and Errors for All Trials

Description of trials	Score	%	Probability	Error description
Arabic numerals, every numeral–numeral trial (*number, not color* error included)	63/84	75	<.01	Said number, not color (11×) Said "none" (4×) Yellow–orange confound (3×) Wrong choice (2×) Said color not on tray
Arabic numerals, every numeral–numeral trial (*number, not color* error not included)	74/84	88.1	<.001	(See above)
Arabic numerals, differing by 1 unit	14/20	70	.06	Said "none" (3×) Yellow–orange confound (2×) Wrong choice
Arabic numerals, differing by 2 units	15/16	93.8	<.01	Yellow–orange confound
Arabic numerals, differing by 4 or 5 units	10/12	83.3	.02	Said "none" Wrong choice
Arabic numerals, same value, same size	10/12	83.3	.02	Said number Said color not on tray
Arabic numerals, same value, different size	12/12	100	<.001	
Arabic numerals and objects, noncongruent[a]	13/16	81.3	.01	Said color not on tray (2×) Same value, responded with respect to mass
Arabic numerals and one object	0/5	0	NA	Always stated "none"

[a] "Noncongruent" means that the value of the Arabic number is larger than or equal to the value of the physical set of items.

I also compared Alex's responses for small (1:2) versus large (5:6) numerals to see if he had more difficulty distinguishing differences between larger numerals. There were, of course, only four trials per pair, too few for meaningful statistical analysis.[7] Nevertheless, the types of errors are interesting. For 1:2, he was correct only 50%; of his two errors, one was a response of "none" and the other was to give the color of the bigger, rather than the smaller, numeral. For 5:6, he was again correct only 50%, but of his two errors, one was a yellow– orange color confound and the other was to provide the number rather than the color.

Identical Numerals

Alex had two sets of this type of question: one set in which the Arabic numbers were the same value and the same size, and one in which the Arabic numbers were the same value but different sizes (see Table 1). On the same value–same size trials, his accuracy was 10/12, or 83.3%, $p < .01$ (binomial test, chance of 1/3; answers could be one of the two colors or "none"); for all trials his result was 12/15, or 80%. On

one trial, he produced and then repeated the label for the number; on another trial he produced the label for a color that was not on the tray. For the same value–different size trials, counting as correct either "none" or the color label of the targeted number, his accuracy was 12/12, or 100%, $p < .01$ (binomial test, chance of 2/3, a color or "none"). Seven times he gave the correct color, five times he said "none," but he gave colors most often in the earlier trials and "none" most often in later trials (see Discussion).

Symbol Versus Collection Trials

Overall, Alex scored 17/26, or 65.4%, on first trials, $p = .08$ (binomial test, chance of 1/2); for all trials, he scored 26/35, or 74.9%; his results were not statistically significantly above chance, although the trend was in the appropriate direction (Table 1 and the Appendix). Reanalyzing the results somewhat differently, however, demonstrates a more complex situation. On trials in which an array of objects was paired with a single Arabic number representing a quantity larger than or equal to the array, but *not* pitting one object against a single numeral representing a larger number, Alex scored 13/16, or 81%, $p = .01$; of the three errors, two involved production of a color not on the tray. Adding the trials in which the single Arabic number represented a quantity less than the array (congruent trials) gives an accuracy of 16/21, or 76.2%, $p = .01$. The large number of errors and lack of statistical significance for the total number of trials arose from five of five trials in which a *single* object was paired with a *single* Arabic number that represented a larger quantity; Alex consistently replied "none." The right versus left position of the smaller number did not make a difference on any of the trials, as errors were random (of nine errors, four were on the right and five on the left).

Interobserver Reliability

The two evaluators who could hear all six trials agreed 100% with what Alex said. The one evaluator who could hear only one trial identified that label correctly. The one evaluator who could hear three trials identified one correctly and two incorrectly. Thus the sum was 13/15, or 86.7%.

Discussion

Despite never having been taught the correspondence between physical sets of objects and Arabic numerals, Alex demonstrated knowledge of ordinality and an equivalence relationship. He also responded in intriguing ways when shown numerals of equal value but different sizes and when collections of objects were contrasted with numerals. In all cases, three serial processing stages were likely to have been involved: (a) identification (recognizing the numerals or quantity of objects), (b) comparison, and (c) response (Dehaene, 1996). The types of processing with respect to comparison might, however, vary depending on the task.

Ordinality

Alex demonstrated that he inferred the ordinal relations among his Arabic numbers without direct training of the 1:1 correspondence between these numerals and their corresponding object sets and without having been trained to recite the numerals in order. Even when errors such as color confounds and providing the correct number rather than color label are included—the latter more a failure to follow directions rather than a failure of numerical understanding— his score is statistically significant. Given that Alex did not receive corresponding amounts of food for his answers (e.g., Olthof & Roberts, 2000; Washburn & Rumbaugh, 1991), he was not responding on the basis of directly associated physical mass, and, given that the numerals themselves were (except for the separate subset of trials) roughly the same size (i.e., granted "1" is of less mass and contour than "6," but their heights are equivalent, and other Arabic numerals, e.g., "2" and "5" are fairly equivalent), he was not likely responding on the basis of the mass or contour of the numerals. In fact, his error for a 1:2 discrimination was, when asked about the

smaller numeral, to give the color of the bigger. He did, for the most part, respond with respect to the numerical magnitude represented by the Arabic numeral. He succeeded on this task by equivalence, that is, by knowing the relationships between the set of items (abstract numerical magnitude) and the English vocalization, the English vocalization and the Arabic numbers, and inferring the relationship between the numerical magnitude and the Arabic number. Little if any learning occurred during testing. Although the few errors he made initially demonstrated a lack of attention to the question and the few errors he made at the end appeared not to be numerically related, his overall scores for initial trials were not significantly different from scores for final trials. Alex's errors showed that he responded in ways both similar to and different from humans when faced with symbolic ordinality tasks.

He showed a limited numerical distance effect, in which discrimination between two numbers decreases as numerical distance between them decreases (Moyer & Landauer, 1967), but not much of a number size effect, in which discrimination of two numbers declines as their numerical size increases, even if the numerical distance remains the same (e.g., discriminating 5:6 is more difficult than 1:2; see Rousselle, Palmer, & Noël, 2004). Of his overall six errors (not including stating the number or color confounds), four were to say "none" for 1:2, 2:3, 4:5, and 1:5 pairs; thus a weak trend was in the direction of more errors for trials involving one unit differences, but for smaller numeral pairs. He seemed to confuse the 1:2 pairing, giving the color of the bigger rather than the smaller numeral once and failing to distinguish between them once; his errors for the 5:6 pairings were a color confound and to give the correct number instead of its requested color, errors not directly related to numerical understanding. For humans, most researchers propose that mental representations become approximate above four, unless canonical patterns (e.g., dice, dominoes) are used (see Mix et al., 2002), and thus comparing mental representations for two large amounts is more difficult than for small amounts. Some researchers argue that for humans numerical distance for two close digits (e.g., 1:2) may influence performance even when the questions do not relate to numerical magnitude and that such distance effects indicate automatic processing of representations of numerical magnitude (note Dehaene & Akhavein, 1995); even training cannot remove the effect in adults (Dehaene, 1997). It is interesting to note that Alex's previous data on vocal discriminations of sets of physical objects that differed by small amounts did not show such an effect (Pepperberg & Gordon, 2005), somewhat like data for children (Yonas, Granrud, & Pettersen, 1985). That the issue arose when Arabic symbols were the stimuli, and once for a larger set (i.e., involving a symbol above "4"; see Buckley & Gillman, 1974), suggests that Alex, like humans, may be converting the numbers to a representation of object quantity, which he then compared mentally, and for which confusion is more likely for comparisons close in size. For children, the ability to classify Arabic numerals as large versus small (and thus make comparisons) improves with age and experience and relies to some extent on memory (see Rubinsten et al., 2002); such was unlikely for Alex. He did not, however, make significantly more errors overall for sets that differed by one unit than for sets that differed by four or five units.

Alex did not appear to be influenced by something called the "spatial–numerical association of response codes" (SNARC) effect. SNARC is based on the idea that a subject orders numerals along a line, right to left with respect to increasing value (based on direction of writing), such that responses would be, in humans, faster in a comparison when the smaller number is to the left and the larger to the right (i.e., ordinal order; Berch, Foley, Hill, & Ryan, 1999; Dehaene et al., 1993, but see Ito & Hatta, 2004). Although Alex's reaction times are not valuable indicators of ability (Pepperberg, 1999), I could determine if he was less accurate when asked to designate the bigger number if it was to the left of the smaller. He was not. Of course, lack of errors does not rule out use of a number line, but given his lack of training on ordinality, he is probably less likely than children to use a number line.

Symbolic Versus Physical Relations

In the two different types of trials in which size was contrasted with numerical quantity, Alex reacted in some ways similar to those of young children (Girelli et al., 2000; Rubinsten et al., 2002). When the sizes of the numerals were different, he initially responded with respect to the physical dimension. When a numeral was contrasted with a physical grouping, he mostly responded on the basis of the numeral quantity; his one exception was quite intriguing. I discuss the different results in turn.

Size of numerals. Although data for children in trials involving the size of numerals are based on reaction times, which cannot be determined for Alex, his response patterns can be used to suggest parallels with human processing. Young children without much experience in comparing numerical values of Arabic digits take significantly more time to respond on the basis of numerical than physical information in determining relative sizes of Arabic numeral pairs when instructed to choose on the basis of numerical value (Rubinsten et al., 2002); Alex initially chose to respond on the basis of the physical relations when number and size were equally valid attributes. Only toward the end of testing did he begin to respond with respect to the symbolic values. Thus, for the first six of these trials, physical comparison, based on an obvious dimension, seems to have interfered with the semantic processing in five cases (see Rubinsten et al., 2002; note also Piaget, 1952). Of the last six trials, in the final four he responded "none"—that is, on the basis of the symbolic value. Given that he was rewarded for both responses, he may simply have begun to change the basis for his answers as testing proceeded and he acquired more experience with the overall task, which emphasized the symbolic value of the numerals. Such a finding suggests that for Alex, like young children, the ability to respond on the basis of relative physical size likely emerges early in development and is somewhat automatic (Pepperberg & Brezinsky, 1991; Yonas et al., 1985), whereas retrieval of relative numerical representations (e.g., a number line, an object file, some mental canonical set) depends on memory and experience. Remember that Alex has not been shown the Arabic numerals together with the physical sets representing those quantities. Thus it was not entirely surprising that Alex, basically like young children who lack experience, initially chose the simpler comparison (see discussion in Girelli et al., 2000). Note that children in these experiments were also given fully incongruent trials (e.g., 2:$_5$), which Alex did not experience. How he would perform on such a difficult task is the subject of future study.

Numerals and physical objects. For the majority of trials, Alex correctly responded to the question of "What color number bigger/smaller?" whether the targeted response was the numeral or the set of objects. Although his overall score was not quite statistically significant (p = .08), the type of errors suggests that, with one exception, he did master the task. His one striking form of error, on comparisons between any Arabic number and a single object—his consistent response of "none"—suggests that his semantic understanding can be overridden when *nothing* physically appears bigger or smaller. It is not possible to determine exactly what cognitive processes were occurring, but the likelihood is that, as in the comparisons between differently sized numerals, the simplest, somewhat automatic response (i.e., one based on physical size) overrode processes involving memory and experience when comparisons involved a perceptual confound of this nature. His other errors involved producing a color not present on the tray, which suggests inattention (see discussion in Pepperberg & Gordon, 2005), and, in one of six instances in which the array and the numeral represented the same quantity, he responded with the color of the collection for the larger mass. Note that this particular trial was the very last of the experiment, and both Alex's attention and motivation were flagging (Pepperberg & Gordon, 2005). Overall, the data suggest that Alex does indeed understand the abstract nature of symbols and that he distinguishes mass and number (see Mix et al., 2002; Sophian, 2000).

Also of interest were the few trials in which the value of the Arabic number and the array of physical items were not pitted against each other, that is, were congruent. In the first such trial, I used five candy hearts

versus the Arabic No. 3; when asked "What color number smaller?," Alex first said "five"—the number of the much more desirable candies, but not, as queried, their color.

Although Alex could ask for anything he wished as a reward, including the candies, possibly their physical presence interfered with his response, such as occurs with chimpanzees (Boysen & Berntson, 1995). Alex was, however, correct on the second try—he produced the appropriate color—and also on subsequent trials of this form, even those involving favored jelly beans.

Equivalence Relations

The data show that Alex—a nonmammal—demonstrated emergence of a strong equivalence relationship. He had never been trained to associate Arabic symbols with the relevant collections of physical objects (see discussions in Dube, McIlvane, Callahan, & Stoddard, 1993; Harrison & Green, 1990; Navarick, Bernstein, & Fantino, 1990), nor had he been trained with either the Arabic symbols or vocal labels in numerical order. Nevertheless, ordinality emerged as a consequence of his training to associate both Arabic numerals and object aggregates with the same vocal label.

Equivalence ties back to the relationship between language and numerical ability. For some researchers, equivalence is the root of symbolic function and a key feature of language (Hineline, 1997; Sidman, 1990; Zentall, 2000; note Hayes, Barnes-Holmes, & Roche, 2001); others consider the reverse to be true (Horne & Lowe, 1996; Lowe & Horne, 1996; see also Dugdale & Lowe, 1990). No one has, however, conclusively produced evidence for either argument (note, however, Hayes et al., 2001).

The issue likely arises because different types of equivalence exist, which involve different types of behavior (Wirth & Chase, 2002) and which may involve different cognitive processes.[8] One form, which I label *weak* with respect to its connection to language, is termed functional equivalence, and requires a common group of stimuli to control a behavior (Sidman, 1990; Zentall, Clement, & Weaver, 2003). Thus, for example, Stimuli A and B both activate a choice of red, whereas Stimuli D and E activate a choice of green, and if C is associated with A and B, it will also likely activate a choice of red; likewise for F, D, E, and green. In humans, what I term *strong* equivalence class formation (i.e., stimulus, not functional, equivalence) with respect to its relation to language requires showing that an item and its written and vocal labels are fully interchangeable (i.e., requires cross-modal transfer; note Sidman, Wynne, Maguire, & Barnes, 1989; but see Devany, Hayes, & Nelson, 1986) and the ability to accept multiple representations for an item. Thus not only would the written label *car* and a toy car elicit the label "car," but hearing "car" would elicit choice of either, and *car* would also elicit choice of the toy and vice versa. If a nonmammal with a very limited language-based code but with clear symbol use shows strong equivalence, full-blown language is unnecessary for such equivalence (as is a mammalian brain). Devising a parallel animal task to test for this form of stimulus equivalence, particularly one including cross-modal transfer, is difficult; but cross-modal transfer ensures that the task is as complex as that for humans. Sea lions (*Zalohus californianus*) relate multiple, disparate icons, but not across modalities (Kastak, Schusterman, & Kastak, 2001; Reichmuth & Schusterman, 2002; Schusterman & Kastak, 1993, 1998), and dolphins (*Tursiops truncatus*) show functional equivalence solely within the auditory mode (von Fersen & Delius, 2000). Some language-trained chimpanzees seem to have problems even with functional symmetry (i.e., comprehension of a symbol they can use productively; Dugdale & Lowe, 2000), although the result seems to be an effect of their initial training, which was primarily use of computer-based lexigrams to request objects (Savage-Rumbaugh, Rumbaugh, & Boysen, 1980; but see Savage-Rumbaugh et al., 1993). Chimpanzees trained on America Sign Language (ASL) and who heard English in their surroundings were taught to associate ASL signs with speech for objects routinely labeled in English (but not shown during training) and, without further training, labeled the objects in ASL—a cross-modal task (Fouts, Chown, & Goodin, 1976). Pigeons seem not to grasp stimulus equivalence (Lionello-

DeNolf & Urcuioli, 2002) but may show functional equivalence (e.g., Jitsumori, Sieman, Lehr, & Delius, 2002; Urcuioli et al., 2002), as do rats (*Rattus norvegicus*; e.g., Nakagawa, 1999). Note that the human task avoids reinforcers that are biologically significant events—that is, appearance or delay of food—which may influence outcomes for animals (review in Astley, Peissig, & Wasserman, 2001).

Alex's results demonstrate that strong stimulus equivalence can exist in the absence of a communication system isomorphic with human language. What does appear necessary for strong (as compared with weak) equivalence is the ability to produce and comprehend symbols—to be able to form and manipulate representations to which the symbols refer (note Devany et al., 1986). For Alex, the numerical values of the Arabic numerals were learned by virtue of their sharing the same *vocal labels* as specific object aggregates, not with respect to any shared physical attributes; that is, they were learned by virtue of a symbolic representation (Wilkinson & McIlvane, 2001). Each item—the Arabic numeral, the spoken label, the aggregate collection—had come to stand for the same thing. Thus a parrot who already demonstrated emergent symmetry (vocally labels a given item and, without training, indicates the item [e.g., by color] when given its vocal label; Pepperberg, 1992) exhibited strong equivalence. Unlike tasks presented to other animal subjects (e.g., sea lions) that did not enable testing across modalities, Alex's task, somewhat like that presented to ASL-trained chimpanzees (Fouts et al., 1976), required the same level of processing as that presented to humans (Devany et al., 1986). Note, of course, that the vocal mode itself is not critical; a chimpanzee trained solely in ASL is also a candidate for such a study. My point is that a connection likely exists between symbolic skills and strong equivalence, and between symbolic skills and numerical competence, and that animals well trained in a symbolic code are likely to provide intermediary links between animals that lack such training and children.

Brain Mechanisms

Alex's data do not, however, provide information about the brain mechanisms that a parrot might use to perform these tasks. Some researchers (e.g., Jarvis et al., 2005) have argued that, for vocal learning, avian and human brains may be more homologous than analogous. Might the same be true for numerical abilities?

Some evidence exists that, at least for humans, specific brain areas are involved in representations of numerical quantities (e.g., Dehaene et al., 2003). For example, the human inferior parietal cortices appear responsible for abstract representation of Arabic numerals (Temple & Posner, 1998); the intraparietal sulcus (IPS) and inferior parietal lobule are supposedly tied to numerical competence (e.g., Lemer, Dehaene, Spelke, & Cohen, 2003) but are distinct from some classical language areas (Cohen, Dehaene, Chochon, Lehe´ricy, & Naccache, 2000). Additional data (e.g., Chochon, Cohen, van de Moortele, & Dehaene, 1999) suggest that only one human brain area, the right postcentral sulcus, is more active in number comparison (i.e., internal manipulation of quantity) than in digit naming, whereas Rickard et al. (2000) seem to find *bilateral* inferior parietal cortical involvement, with only a bias toward the right. Rickard et al. actually suggested that significant individual differences may exist for magnitude comparisons, possibly because such processing is highly distributed.

Could the parrot brain be similarly organized for numerical competence? Note, however, that most human number comparison tasks involve issues of spatial attention and nonsymbolic comparisons (assuming that the Arabic numbers recall a representation of physical quantity and not simply a memorized number string), which are also correlated to IPS activity (Coull & Nobre, 1998; Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Göbel et al., 2004; Jordan, Schadow, Wuestenberg, Jeinze, & Jäncke, 2004; Simon, 1999; but see Spelke & Dehaene, 1999).[9] As noted previously (Pepperberg, 2006), such abilities are essential for nonhuman survival (e.g., for predator detection, for foraging comparisons), and nonhumans are thus likely to have analogous, if not homologous, brain areas (note Walsh, 2003). It is

interesting to note that some data (Kawashima et al., 2004) suggest that either maturation or experience might affect how the brain processes numerical information in humans (note Gordon, 2004; but see Temple & Posner, 1998); might the same be true for Alex and other nonhumans given extensive training on numerical symbols (e.g., Biro & Matsuzawa, 2001; Boysen & Berntson, 1989)? Complicating the issue is that Cohen et al. (2000) specifically stated that the ability to decide nonverbally which of two Arabic numbers is larger remains in human patients even with the largest left hemispheric lesions (i.e., when producing a vocal number sequence or other verbal tasks involving numbers are impaired; see also Colvin, Funnell, & Gazzaniga, 2005; Rickard et al., 2000), suggesting some dissociation between vocal and nonvocal abilities that may separate Alex from other number-trained animals.[10] Although no neurobiological number studies have been performed on birds, possibly the avian nidopallium caudolaterale and the dorsolateral corticoid areas, which respectively are likely homologues of the posterior parietal cortex and parts of the prefrontal cortex, are involved in numerical processing (A. Butler, personal communication, July 5, 2005; E. Jarvis, personal communication, July 12, 2005). Thus, as before (Pepperberg, 2006), I suggest that whatever brain areas Alex uses do indeed function in at least an analogous manner, given the similarities in data.

Summary

In sum, Alex demonstrated that ordinality may be an emergent property in a nonhuman, arising somewhat as it does it for small children, even though his training did not replicate the ways in which children are exposed to numerical concepts. In fact, Alex's exposure to and training on numerical tasks differ considerably from those of normal children. Note, however, that humans who lack training may not be as competent as Alex (Gordon, 2004). Alex also demonstrated that a humanlike form of stimulus equivalence is within the realm of a nonhuman, nonprimate, nonmammal that has received some cross-modal symbolic instruction. Such data begin to provide some answers as to the extent to which human and nonhuman numerical concepts may differ and still be comparable.

Notes

[1] Note, however, that some question exists as to whether these brain areas are exclusively involved in numerical competence; see Summary and Discussion in, for example, Göbel, Johansen-Berg, Behrens, and Rushworth (2004).

[2] Because Alex would be rewarded for the test trials to maintain his interest in the task, he still could learn the correct response (which is why the first set of trials will be most important); however, because the response is not the number label and he could be asked about either the larger or the smaller value, the response would not be a simple association of "If I see X and Y, the answer is Y."

[3] Remember that the point of these studies was to see if the ordinal relationship would emerge from the training that had occurred to date; thus additional training was specifically to be avoided.

[4] Thus, if Alex made one error, the total number of trials presented for a given question would be two and his score 1/2; if he made four errors on a given question his score would be 0/4, and I would then go onto another question.

[5] A contextless tape is one from which all questions have been edited, so that the transcriber hears only Alex's responses. The reported percentage represented 106 matches of 108 vocalizations. As an additional control, I made two transcriptions of a student, new to the lab, as he responded to the same type of questions as Alex. The first transcription was live; the second, made several days later, was also

contextless. The two transcriptions of the student's vocalizations matched to within 95.8% (68 of 71 vocalizations).

[6] Wooden numerals are always standardized with nontoxic paints for which Alex has been trained to use a specific color label; however, Alex often chews up these numerals as his reward and they are now difficult to replace. I therefore had to use commercially available magnetized plastic "refrigerator letters" for much of the study, and color errors occurred with these numerals. Paint that would stick to plastic is toxic, and because Alex is given the numeral to chew as his initial reward, it could not be used. Note that he did not always err on such items and would answer correctly when asked a second time, suggesting that the color clearly was on a boundary (if not yellow, it had to be orange). Possibly, too, the error had to do with attention or motivation because on separate trials in which he was just asked the color of the object, he learned the appropriate label.

[7] Remember, the number of trials was deliberately kept low to avoid training.

[8] Note that some researchers would disagree with this statement (see Vaughan, 1988).

[9] Coull and Nobre (1998) suggested that the left IPS is more active in temporal attention and the right IPS in spatial attention, but Fias et al. (2003) found more activity in the left IPS for representation of symbolic and nonsymbolic magnitude. Cohen et al. (2000) showed that small bilateral intraparietal foci are more active in calculation abilities than in control tasks that also involve similar types of spatial or attentional abilities. All cited studies do, however, implicate the IPS in numerical tasks.

[10] And these issues are complicated even more because studies on brain hemispheres and specific brain areas suggest activation may differ for values ≤4 or >4 (Colvin et al., 2005; see also Göbel, Rushworth, & Walsh, 2001; Göbel, Walsh, & Rushworth, 2001).

References

Astley, S. L., Peissig, J. J., & Wasserman, E. A. (2001). Superordinate categorization via learned stimulus equivalence: Quantity of reinforcement, hedonic value, and the nature of the mediator. *Journal of Experimental Psychology: Animal Behavior Processes, 271,* 252–268.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology, 74,* 286–308.

Biro, D., & Matsuzawa, T. (2001). Use of numerical symbols by the chimpanzee (*Pan troglodytes*): Cardinals, ordinals, and the introduction of zero. *Animal Cognition, 4,* 193–199.

Bowmaker, J. K., Heath, L. A., Das, D., & Hunt, D. M. (1994). Spectral sensitivity and opsin structure of avian rod and cone visual pigments. *Investigative Ophthalmology and Visual Science, 35,* 1708.

Bowmaker, J. K., Heath, L. A., Wilkie, S. E., Das, D., & Hunt, D. M. (1996). Middle-wave cone and rod visual pigments in birds: Spectral sensitivity and opsin structure. *Investigative Ophthalmology & Visual Science, 37,* S804.

Boysen, S. T. (1993). Counting in chimpanzees: Nonhuman principles and emergent properties of number. In S. T. Boysen & E. J. Capaldi (Eds.), *The development of numerical competence: Animal and human models* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 103,* 23–31.

Boysen, S. T., & Berntson, G. G. (1990). The development of numerical skills in the chimpanzee. In S. T. Parker & K. R. Gibson (Eds.), *"Language" and intelligence in monkeys and apes: Comparative developmental perspectives* (pp. 435–450). Cambridge, England: Cambridge University Press.

Boysen, S. T., & Berntson, G. G. (1995). Response to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 82–86.

Boysen, S. T., Berntson, G. G., Shreyer, T. A., & Quigley, K. S. (1993). Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 107,* 208–215.

Boysen, S. T., & Hallberg, K. I. (2000). Primate numerical competence: Contributions toward understanding nonhuman cognition. *Cognitive Science, 24,* 423–443.

Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition, 83,* 223–240.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1–9 by monkeys. *Science, 282,* 746–749.

Brannon, E. M., & Terrace, H. S. (2000). Representation of the numerosities 1–9 by rhesus macaques. *Journal of Experimental Psychology: Animal Behavior Processes, 26,* 31–49.

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science, 12,* 238–243.

Bruce, B., & Threfall, J. (2004). One, two, three and counting. *Educational Studies in Mathematics, 55,* 3–26.

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136.

Carey, S. (2004, Winter). Bootstrapping & the origin of concepts. *Dædalus,* 1–10.

Chochon, F., Cohen, L., van de Moortele, P. F., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience, 11,* 617–630.

Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., & Naccache, L. (2000). Language and calculation within the parietal lobe: A combined cognitive, anatomical, and fMRI study. *Neuropsychologia, 38,* 1426–1440.

Colvin, M. K., Funnell, M. G., & Gazzaniga, M. S. (2005). Numerical processing in the two hemispheres: Studies of a split-brain patient. *Brain & Cognition, 57,* 43–52.

Coull, J. T., & Nobre, A. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience, 18,* 7426–7435.

Davis, H., & Pérusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11,* 561–615.

Dehaene, S. (1996). The organization of brain activation in number comparison: Event related potentials and the additive-factors method. *Journal of Cognitive Neuroscience, 8,* 47–68.

Dehaene, S. (1997). *The number sense.* Oxford, England: Oxford University Press.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 314–326.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371–396.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology, 14,* 218–224.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20,* 487–506.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46,* 243–257.

Dube, W. V., McIlvane, J., Callahan, T. D., & Stoddard, L. T. (1993). The search for stimulus equivalence in nonverbal organisms. *Psychological Record, 43,* 761–778.

Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackmans & H. Lejune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 115–138). Hove, England: Erlbaum.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73,* 5–22.

Emmerton, J., Lohmann, A., & Niemann, J. (1997). Pigeons' serial ordering of numerosity with visual arrays. *Animal Learning & Behavior, 25,* 234–244.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience, 15,* 47–57.

Fouts, R. S., Chown, B., & Goodin, L. (1976). Transfer of signed responses in American Sign Language from vocal English stimuli to physical object stimuli by a chimpanzee (*Pan*). *Learning and Motivation, 7,* 458–475.

Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer-Verlag.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43–74.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77,* 236–263.

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number* (2nd ed.). Cambridge, MA: Harvard University Press.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology, 76,* 104–122.

Göbel, S. M., Johansen-Berg, H., Behrens, T., & Rushworth, M. F. S. (2004). Response-selection-related parietal activation during number comparison. *Journal of Cognitive Neuroscience, 16,* 1536–1551.

Göbel, S. M., & Rushworth, M. F. S. (2004). Cognitive neuroscience: Acting on numbers. *Current Biology, 14,* R517–R519.

Göbel, S., Rushworth, M. F. S., & Walsh, V. (2001). rTMS disrupts the representation of small numbers in supramarginal gyrus [Abstract]. *NeuroImage, 13*(6, Suppl. 1), 409.

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001). The mental number line and the human angular gyrus. *NeuroImage, 14,* 1278–1289.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306,* 496–499.

Harrison, R. J., & Green, G. (1990). Development of conditional and equivalence relations without differential consequences. *Journal of the Experimental Analysis of Behavior, 54,* 225–237.

Hayes, S., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum Press.

Hineline, P. N. (1997). How, then, shall we characterize this elephant? *Journal of the Experimental Analysis of Behavior, 68,* 297–300.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185–241.

Ito, K., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition, 32,* 662–673.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate evolution. *Nature Reviews Neuroscience, 6,* 151–159.

Jitsumori, M., Sieman, M., Lehr, M., & Delius, J. D. (2002). A new approach to the formation of equivalence classes in pigeons. *Journal of the Experimental Analysis of Behavior, 78,* 397–408.

Jordan, K., Schadow, J., Wuestenberg, T., Jeinze, H.-J., & Jäncke, L. (2004). Different cortical activations for subjects using allocentric or egocentric strategies in a virtual navigation task. *Brain Imaging, 15,* 135–140.

Judge, P. G., Evans, T. A., & Vyas, D. K. (2005). Ordinal representation of numeric quantities by brown capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 79–94.

Kastak, C. R., Schusterman, R. J., & Kastak, D. (2001). Equivalence classification by California sea lions using class-specific reinforcers. *Journal of the Experimental Analysis of Behavior, 76,* 131–158.

Kawai, N., & Matsuzawa, T. (2000). Numerical memory span in a chimpanzee. *Nature, 403,* 39–40.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., et al. (2004). A functional MRI study of simple arithmetic: A comparison between children and adults. *Cognitive Brain Research, 18,* 225–231.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia, 41,* 1942–1958.

Liddle, I., & Wilkinson, J. E. (1987). The emergence of order and class aspects of number in children: Some findings from a longitudinal study. *British Journal of Educational Psychology, 57,* 237–243.

Lionello-DeNolf, K. M., & Urcuioli, P. J. (2002). Stimulus control topographies and test of symmetry in pigeons. *Journal of the Experimental Analysis of Behavior, 78,* 467–495.

Lowe, C. F., & Horne, P. J. (1996). Reflections on naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 315–340, 341–353.

Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315,* 57–59.

Matsuzawa, T., Itakura, S., & Tomonaga, M. (1991). Use of numbers by a chimpanzee: A further study. In A. Ehara, T. Kimura, O. Takenaka, & M. Iwamoto (Eds.), *Primatology today* (pp. 317–320). Amsterdam: Elsevier.

Mix, K., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood.* New York: Oxford University Press.

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting internal psychology. *Perception & Psychology, 13,* 180–184.

Moyer, R. S., & Landauer, T. K. (1967). The time required for judgements of numerical inequality. *Nature, 215,* 1515–1520.

Nakagawa, E. (1999). Acquired equivalence of discriminative stimuli following two concurrent discrimination learning tasks as a function of overtraining. *Psychological Record, 49,* 327–348.

Navarick, D. J., Bernstein, D. J., & Fantino, E. (1990). The experimental analysis of human behavior. *Journal of the Experimental Analysis of Behavior, 54,* 159–162.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37,* 149–157.

Olthof, A., Iden, C. M., & Roberts, W. A. (1997). Judgments of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 325–333.

Olthof, A., & Roberts, W. A. (2000). Summation of symbols by pigeons (*Columba livia*): The importance of number and mass of reward items. *Journal of Comparative Psychology, 114,* 158–166.

Pepperberg, I. M. (1981). Functional vocalizations by an African Grey parrot. *Zeitschrift für Tierpsychologie, 55,* 139–160.

Pepperberg, I. M. (1987). Evidence for conceptual quantitative abilities in the African Grey parrot: Labeling of cardinal sets. *Ethology, 75,* 37–61.

Pepperberg, I. M. (1990). Cognition in an African Grey parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology, 104,* 41–52.

Pepperberg, I. M. (1992). Proficient performance of a conjunctive, recursive task by an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology, 106,* 295–305.

Pepperberg, I. M. (1994). Evidence for numerical competence in an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology, 108,* 36–44.

Pepperberg, I. M. (1997). *Vocal labeling of Arabic numerals.* Unpublished manuscript.

Pepperberg, I. M. (1999). *The Alex studies: Cognitive and communicative abilities of Grey parrots.* Cambridge, MA: Harvard University Press.

Pepperberg, I. M. (2006). Addition by a Grey parrot (*Psittacus erithacus*), including absence of quantity. *Journal of Comparative Psychology, 120,* 1–11.

Pepperberg, I. M. (in press). Grey parrots do not always "parrot": Roles of imitation and phonological awareness in the creation of new labels from existing vocalizations. *Language Sciences.*

Pepperberg, I. M., & Brezinsky, M. V. (1991). Acquisition of a relative class concept by an African Grey parrot (*Psittacus erithacus*): Discriminations based on relative size. *Journal of Comparative Psychology, 105,* 286–294.

Pepperberg, I. M., & Gordon, J. D. (2005). Number comprehension by a Grey parrot (*Psittacus erithacus*), including a zero-like concept. *Journal of Comparative Psychology, 119,* 197–209.

Pepperberg, I. M., & Kozak, F. A. (1986). Object permanence in the African Grey parrot (*Psittacus erithacus*). *Animal Learning & Behavior, 14,* 322–330.

Pepperberg, I. M., & Wilkes, S. R. (2004). Lack of referential vocal learning from LCD video by Grey parrots (*Psittacus erithacus*). *Interaction Studies, 5,* 75–97.

Piaget, J. (1952). *The child's concept of number.* New York: Norton.

Premack, D. (1976). *Intelligence in ape and man.* Hillsdale, NJ: Erlbaum.

Reichmuth, C., & Schusterman, R. J. (2002). Sea lions and equivalence: Expanding classes by exclusion. *Journal of the Experimental Analysis of Behavior, 78,* 449–465.

Rickard, T. C., Romero, S. G., Basso, G., Wharton, C., Flitman, S., & Grafman, J. (2000). The calculating brain: An fMRI study. *Neuropsychologia, 38,* 325–335.

Rousselle, L., Palmer, E., & Noël, M. -P. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. *Journal of Experimental Child Psychology, 87,* 57–84.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev. S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology, 81,* 74–92.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 233,* 1–258.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., & Boysen, S. (1980). Do apes use language? *American Scientist, 68,* 49–61.

Schusterman, R. J., & Kastak, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *Psychological Record, 43,* 823–839.

Schusterman, R. J., & Kastak, D. (1998). Functional equivalence in a California sea lion: Relevance to animal social and communicative interactions. *Animal Behaviour, 55,* 1087–1095.

Sidman, M. (1990). Equivalence relations: Where do they come from? In H. Lejeune & D. Blackman (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Erlbaum.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

Simon, T. J. (1999). The foundations of numerical thinking in a brain without numbers. *Trends in Cognitive Science, 3,* 363–365.

Sophian, C. (1995). *Children's numbers,* Madison, WI: Brown & Benchmark.

Sophian, C. (2000). From objects to quantities: Developments in preschool children's judgments about aggregate amount. *Developmental Psychology, 36,* 724–730.

Spelke, E. S., & Dehaene, S. (1999). Biological foundations of numerical thinking. *Trends in Cognitive Science, 3,* 365–366.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78,* 45–88.

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-olds and adults. *Proceedings of the National Academy of Sciences, USA, 95,* 7836–7841.

Teubal, E., & Guberman, A. (2002). The development of children's counting ability. *Megamot, 42,* 83–102.

Thomas, R. K., Fowlkes, D., & Vickery, J. D. (1980). Conceptual numerousness judgments by squirrel monkeys. *American Journal of Psychology, 93,* 247–257.

Tomonaga, M., Matsuzawa, T., Itakura, S. (1993). Teaching ordinals to a cardinal trained chimpanzee. *Primate Research, 9,* 67–77.

Urcuioli, P. J., Pierce, J. N., Lionello-DeNolf, K. M., Friedrich, A., Fetterman, J. G., & Green, C. (2002). The development of emergent differential sample behavior in pigeons. *Journal of the Experimental Analysis of Behavior, 78,* 409–432.

Vaughan, W., Jr. (1988). Formation of equivalence sets in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 36–42.

von Fersen, L., & Delius, J. D. (2000). Acquired equivalences between auditory stimuli in dolphins (*Tursiops truncatus*). *Animal Cognition, 3,* 79–83.

Walsh, V. (2003). Cognitive neuroscience: Numerate neurons. *Current Biology, 13,* R447–R448.

Washburn, D. A., & Rumbaugh, D. A. (1991). Ordinal judgements of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science, 2,* 190–193.

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

Wilkinson, K. M., & McIlvane, W. J. (2001). Methods for studying symbolic behavior and category formation: Contributions of stimulus equivalence research. *Developmental Review, 2,* 355–374.

Wirth, O., & Chase, P. N. (2002). Stability of functional equivalence and stimulus equivalence: Effects of baseline reversals. *Journal of the Experimental Analysis of Behavior, 77,* 29–47.

Wynne, C. D. L. (1992, May). *Animal inferences: Complex performances–simple mechanisms.* Paper presented at the meeting of the Midwest Psychological Association, Chicago.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1–B11.

Yonas, A., Granrud, C. E., & Pettersen, L. (1985). Infants' sensitivity to relative size information for distance. *Developmental Psychology, 21,* 161–167.

Zentall, T. R. (2000). Symbolic representation by pigeons. *Current Directions in Psychological Science, 9,* 118–123.

Zentall, T. R., Clement, T. S., & Weaver, J. E. (2003). Symmetry training in pigeons can produce functional equivalences. *Psychonomic Bulletin & Review, 10,* 387–391.

Appendix

Complete Set of Symbol Versus Collection Trials for Objects and Numerals

Object	Numeral	Bigger vs. Smaller	Response[a]	R vs. L
4 O wool pompon	6 G	bigger	G	R
3 Y wood block	4 Ppl	smaller	Y	L
1 B key	6 Gy	bigger	none (2×)	R
5 G candy heart	3 Y	smaller	5, Y	R
1 Ppl wood stick	2 O	bigger	none (2×)	L
2 Gy wood ring	5 B	smaller	Gy	R
1 B key	1 Re	smaller	none	L
5 Re wool pompon	6 Y	bigger	Y	L
3 G wood block	5 Re	smaller	B, G	L
3 Ppl jelly bean	1 O	bigger[b]	Ppl	R
2 Gy key	4 B	bigger	B	L
5 B wood block	5 G	smaller	none	R
1 Y chalk	5 G	bigger	none (2×)	L
3 G jelly bean	6 Re	bigger	Re	L
2 B key	2 O	smaller	none	R
4 B wood rod	5 Ppl	smaller	G, B	R
4 O wood block	2 Gy	bigger[b]	O	R
2 Ppl wood ring	3 O	smaller	Ppl	L
2 Gy wood block	1B	bigger[b]	Gy	R
1 B jelly bean	4 Re	bigger	none (2×)	L
6 O wool pompon	6 B	smaller	none	R
2 O key	6 Y	smaller	O	R
1 G wood block	3 Re	bigger	none (2×)	L
3 Re wool pompon	3 Y	bigger	none	L
3 Y jelly bean	2 Re	smaller[b]	Re	L
4 B plastic spool	4 O	bigger	B, none	R

Note. Last column states whether the smaller value was to Alex's right (R) or left (L). B = blue; G = green; Gy = grayish-black; O = orange; Ppl = purple; Re = red (Alex's label is "rose"; Pepperberg, 1981); Y = yellow.

[a] If Alex said "none" when there was one object and one numeral, the response was reported as an error, and he was queried only a second time, not four times.

[b] Designates trials in which the object array and Arabic numeral are consistent with the question posed.