

WellBeing International

## WBI Studies Repository

3-1994

# Numerical Competence in an African Gray Parrot (*Psittacus erithacus*)

Irene M. Pepperberg

Follow this and additional works at: [https://www.wellbeingintludiesrepository.org/acwp\\_asie](https://www.wellbeingintludiesrepository.org/acwp_asie)



Part of the [Animal Studies Commons](#), [Comparative Psychology Commons](#), and the [Other Animal Sciences Commons](#)

---

### Recommended Citation

Pepperberg, I. M. (1994). Numerical competence in an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 108(1), 36.

This material is brought to you for free and open access by WellBeing International. It has been accepted for inclusion by an authorized administrator of the WBI Studies Repository. For more information, please contact [wbisr-info@wellbeingintl.org](mailto:wbisr-info@wellbeingintl.org).



# Numerical Competence in an African Gray Parrot (*Psittacus erithacus*)

Irene M. Pepperberg

## ABSTRACT

*An African gray parrot (Psittacus erithacus), Alex, trained to label vocally collections of 1-6 simultaneously presented homogeneous objects, correctly identified, without further training, quantities of targeted subsets in heterogeneous collections. For each test trial Alex was shown different collections of 4 groups of items that varied in 2 colors and 2 object categories (e.g., blue and red keys and trucks) and was asked to label the number of items uniquely defined by the conjunction of 1 color and 1 object category (e.g., "How many blue key?"). The collections were designed to provide maximal confounds (or distractions) and thus replicate the work of Trick and Pylyshyn (1989) on humans. Humans count rather than subitize under such conditions. Alex's results (83.3% overall accuracy) are therefore discussed in terms of their relation to human numerical competence, particularly with respect to counting.*

Several studies have shown that a wide range of vertebrates recognize numerical quantities (e.g., reviews in Boysen & Capaldi, 1992; Davis & Perusse, 1988). Such studies are also often proposed as a basis for cross-species comparisons of general cognitive processing abilities (see Case, 1985; Gelman & Gallistel, 1986; Lenneberg, 1971; cf. Starkey, Spelke, & Gelman, 1990). The rationale is that numerical competence indicates an ability to generalize abstract categories across domains: Number is not an inherent attribute of an object, as is color, shape, or material, but is rather a descriptor that is applicable to any discrete collection of entities. Whether numerical competence indicates general abstract cognitive processing abilities remains, however, to be fully tested.

Central to this issue is what constitutes numerical competence. Unfortunately, for quantities most often studied (1-8), two enumeration processes appear to exist: *Subitizing*, a supposedly fast, effortless, perceptual apprehension of numbers usually up to 4 that uses preattentive mechanisms (Kaufman, Lord, Reese, & Volkman, 1949; Stevens, 1951; Taves, 1941; Wolters, van Kempen, & Wijnhuizen, 1987), and *counting*, a slow, effortful process generally for numbers greater than 4 that requires spatial attention (Aoki, 1977; Atkinson, Campbell, & Francis, 1976; Oyama, Kikuchi, & Ichihara, 1981). Note, however, that some researchers argue that counting does not begin until 6-7 (Mandler & Sheba, 1982; Miller, 1956; see Trick & Pylyshyn, 1989, 1991). During subitizing, number loses its abstract nature as a descriptor to become a holistic attribute, like color or shape (Pepperberg, 1988c). In contrast, counting is seen as a set of abstract steps: One (a) produces a standard sequence of number tags, (b) applies a unique number tag to each item to be counted, (c) remembers what already has been counted, and (d) knows that the last number tag used tells how many objects are there (Fuson & Hall, 1983). Despite arguments to the contrary (e.g., Beckmann, 1924; Gelman & Gallistel, 1986; Mandler & Sheba, 1982), subitizing is

generally viewed as the simpler process (see Cole & Scribner, 1974; Davis & Perusse, 1988, and commentaries). Too, the cognitive overlap of subitizing and counting proposed by Gallistel and Gelman (1991, 1992) seems countered by data that has shown that the overlap is a form of subitizing (Davis & Perusse, 1988), that is, a post-counting process, distinct from precounting perceptual recognition, that through practice allows exceptionally speedy labeling of fewer than 4-6 items. This division retains counting as the purported process for larger quantities.

Whether different processes are used to enumerate small versus large quantities and the relative complexity of such processes must be known if numerical competence can be used to compare cognitive abilities across species. Precounting subitizing, for example, does not require subjects to handle abstract entities, an ability that may indicate more general abstract processing capacities. Unfortunately, even studies in which animals selected or labeled quantity used collections within subitizing range ( $\leq 6$ ; e.g., Boysen & Berntson, 1989; Davis, 1984; Matsuzawa, 1985; Pepperberg, 1987a). Questions thus still exist as to whether (a) processes used by animals to recognize quantities up to 8 involve anything more than simple perception (e.g., Mandler & Shebo, 1982), (b) animals can use for larger amounts the same enumeration process in which adult humans ultimately obtain proficiency, and (c) overall cognitive competence can be judged by numerical competency on small numbers if a simple perceptual mechanism can be used to recognize smaller collections. Clearly, animals that are to be compared with humans must be tested so as to demonstrate the more complex counting procedure.

So far, few nonhumans have met the criteria for human counting. Some data on nonhumans suggest a capacity for counting (e.g., rats, Capaldi & Miller, 1988; cf. Davis & Albert, 1986), but only chimpanzees (*Pan troglodytes*) have shown abilities similar to young children (Boysen & Berntson, 1990). A chimpanzee, Sheba, demonstrated ordinality (Boysen & Berntson, 1988) and labeled, with a card depicting an Arabic numeral, the sum of two arrays separated in time and space (Boysen & Berntson, 1989). She was, however, tested only on sets up to 4. Might other animals show such capacities-or more advanced ones-with proper training?

If capacities necessary for numerical competence are used for other complex cognitive tasks, then the answer is likely affirmative. Specifically, an African gray parrot (*Psittacus erithacus*) that can categorize, comprehend concepts of same-different, absence of information, and relative size, and show certain stages of numerical competence (Pepperberg, 1983, 1987a, 1987b, 1988a; Pepperberg & Brezinsky, 1991) will be a good candidate for such a study. This bird, Alex, can already vocally label collections of 2-6 simultaneously presented homogeneous or heterogeneous (e.g., corks and keys) objects (Pepperberg, 1987b). Tests have ruled out the possibility that the parrot uses cues, such as mass, brightness, surface area, odor, object familiarity, or canonical pattern recognition, but have not ruled out his use of a noncounting strategy, such as *clumping* or *chunking*-a form of subitizing (e.g., perception of 6 as two groups of 3; see Jevons, 1871; Mandler & Shebo, 1982; von Glasersfeld, 1982)-to obtain the correct answer.

### **Choosing a Task to Study Numerical Competence in an Avian Subject**

Demonstrating avian counting behavior is challenging. Birds are extremely sensitive to quantifiable sequential auditory patterns (Seibt, 1982) and, unlike some species (e.g., rats; Davis & Albert, 1987), may transfer this ability to visual patterns. Homogeneous sets of external events may thus simply be judged as to whether they match or fill up specific perceptual patterns (Meek & Church, 1983), even for quantities that humans count. Crows (*Corvus brachyrhynchos*), for example, likely use various numbers of caws with different temporal patterning to identify each other in flocks (Thompson, 1968, 1969); each crow thus judges several different additive acoustic patterns at least to some degree with respect to quantity. Other avian species may recognize particular sets of repetitions of different neighbors' vocalizations so as to

respond appropriately (European blackbirds, *Turdus merula*; Wolfgramm & Todt, 1982; wood peewees, *Contopus virens*, Smith, 1988; carduelid finches and their hybrids, Giittinger, 1979). Such behavior is not unlike human determination, without necessarily counting, of how many times to repeat *fa la la* in a familiar Christmas carol. Birds, however, seem particularly likely to memorize many such perceptual sets. This behavior is often termed *serial subitizing* (Burns, 1988, p. 581; note Davis & Perusse, 1988). Given the possibility of avian auditory-to-visual transfer, the sequential addition process used for the chimpanzee Sheba (Boysen & Berntson, 1990) may not be valid for testing a bird.

An inferential test to distinguish perceptual recognition from counting has, however, been suggested by recent work on humans (Trick & Pylyshyn, 1989). The system is based on visual processing mechanisms and involves distractors (Trick & Pylyshyn, 1991). Subjects enumerate items in a field of distractors under two conditions, (a) either white or vertical lines among green horizontals or (b) white vertical lines among both green verticals and white horizontals. Subitizing was evident for 1-3 only for the first condition. Subitizing thus seemed to fail when visual items required conjunctive attentive processing for recognition, that is, when subjects must distinguish among various objects defined by a collection of features (e.g., color and shape). Such findings are consistent with the suggestion (Glanville & Dallenbach, 1929) that the number of items that can be apprehended simultaneously decreases as the amount to be perceived about them increases. Similar tests can be given to the parrot Alex because he can already use a conjunctive condition to identify a single object within a collection (e.g., a red key within a collection of colored keys and other red items; Pepperberg, 1992). He can now be asked to label the quantity of a subset of items similarly defined. Success may demonstrate that Alex's competency, if not necessarily his strategy, is equivalent to that of humans.

## Method

### *Experimental Design*

*Subject.* Alex, an African gray parrot (*Psittacus erithacus*), resides in a laboratory (details in Pepperberg, 1981); he has been the focus of studies on interspecies communication and animal cognition since June 1977.<sup>1</sup> When this experiment began, he could produce vocal (English) labels to respond to questions about attributes of over 100 different objects, including those that varied somewhat from the training exemplars (Pepperberg, 1990a), and could answer questions about abstract categories of color, shape, material, relative size, and quantity (Pepperberg, 1981, 1983, 1987a, 1987b, 1987c, 1988a, 1988b, 1990a; Pepperberg & Brezinsky, 1991). In this study he used his ability to produce labels for quantity (1-6) and to comprehend labels for colors (rose [red], green, purple, blue, yellow, orange, and gray) and objects (wood [tongue depressors, circles, hearts, spools, etc.], paper, chalk, wool [pompons or felt forms], hide [rawhide], rock [Playdoh forms], block, truck [toy cars of rubber, metal, or plastic], box [metal, paper, or plastic], cup [plastic or paper], chain [of paper clips], and key [plastic or metal]).

*Task.* The task required the numerical and conjunctive capacities Alex had previously demonstrated. In an earlier numerical study, Alex learned to produce vocal numerical labels for sets of 2-6 objects; labels were not trained in numerical order (Pepperberg, 1987b). Without further training he then responded correctly on 7 of 10 first trials to questions about the quantity of a particular targeted subset of objects within a simple heterogeneous display, for example, the number of corks in a collection of corks and keys. In a prior conjunction study (Pepperberg, 1992), Alex had to provide information about the specific instance of one category of an item in a collection that was uniquely defined by the conjunction of two other categories (e.g., "What material is the [object defined by shape and color]?"). His accuracy was 26 of 34 first trials, or 76.5%.

Now Alex was asked about the quantity of one of four subsets of items that varied with respect to two color and two object categories; items in each subset of a collection shared either color or form with those of the other subsets (e.g., red cups, blue cups, red keys, and blue keys). Seven categories existed for colors and 12 for objects; many object categories, furthermore, contained exemplars of different materials (e.g., cups of paper or plastic). In each trial Alex had to produce vocally the label for the quantity of a subset defined by both color and object category (e.g., answer "How many red key?"; see sample data in the Appendix). Each question thus contained two types of information, the topic under study (e.g., the quantity of the subset vs. the attribute of one item in the collection) and the designation (by conjoining labels for color and object category; e.g., red and key, not red and truck) for the subset that was the target of the search. To respond correctly Alex had to process each type of information without error and then recognize and encode as a vocal label (e.g., "two") the information about quantity. Some or all of this behavior likely occurred as separate steps, by which each step added to the complexity of the task (see Premack, 1983).

Alex's collections included one more set of dis tractors than those presented to human subjects (Trick & Pylyshyn, 1991) to counter the possibility that the nature of his exemplars may have somewhat simplified his task: Whereas Alex had concrete objects, such as colored keys and nails, humans had colored lines of different orientations. Alex thus had to label the appropriate quantity from among four rather than three subsets of objects that shared either color or object labels.

As in all studies with Alex, the protocol differed from ones 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 presented intermittently during training and testing of other unrelated topics also under study (e.g., middleness and phoneme labeling; see Discussion in Pepperberg, 1990a). Alex's responses thus had to be chosen from his entire repertoire (more than 90 vocalizations, including labels for foods and locations) and from among numerous possible topics concerning various exemplars and questions during each session. This design not only increased the complexity of the task, but it prevented several forms of cuing (see Premack, 1976, and *Testing Procedures* below).

### *Training Procedures*

Alex's training on labeling quantities of objects is described in detail in Pepperberg (1987b). No additional training occurred before the present task was administered. For the readers' convenience, a brief summary of the general procedures (which follows Pepperberg & Brezinsky, 1991) is presented here.

Training is conducted by interacting human tutors. Alex watches a pair of humans act as trainer and trainee. The trainer questions the trainee about specific objects or concepts and the trainee models the targeted behavior; the trainee is also Alex's rival for the trainer's attention. The two humans exchange roles and demonstrate referential, contextual use of each targeted vocalization. Correct responses are rewarded either with the exemplars that relate to the skill or task being taught or with the right to request (vocally) an alternative object or action (Pepperberg, 1987b, 1988b). Training on a given concept occurs simultaneously with testing and training on other concepts. Training occurs 2-4 times each week, and sessions vary in length from 5 min to 1 hr, depending on Alex's attention span.

### *Testing Procedures*

*Criterion for beginning testing.* Because the range of numbers that must be tested began at 1, I had to ensure that Alex could label single items. The label *one*, although in his repertoire (Pepperberg, Brese, & Harris, 1991), had never been formally tested. I thus used competence in labeling single objects as the criterion for when to begin testing. Using the protocol in Pepperberg (1987b), my students and I queried

Alex from November 1989 to January 1990 about various single objects in the laboratory ("How many?"). Alex reached criterion, 24 initially correct responses in 30 consecutive questions (80%), after 35 trials.

*General test procedures.* Test sessions of a single numerical question on heterogeneous collections occurred on average 1-2 times each week during January-August 1990, March-April 1991, and June-August 1992.<sup>2</sup> The following is a summary of the rationale for the test procedures (details are in Pepperberg, 1981, 1990a).

Tests were designed to avoid an examiner's expectation of a certain type of answer. An examiner who, for example, poses a series of number-related questions may come to expect numerical answers and unconsciously accept an indistinct (and by our criteria incorrect) response of "gree" (a mix of *green* and *three*) for *three*. Two precautions were taken so that an examiner could not predict which of 17 different questions on 9 possible topics (or about 90 answers) would appear on a given day. First, tests covered all topics Alex had mastered. Thus, for each test the principal trainer listed all possible questions and objects to be tested for all of the topics under examination (e.g., number competency, same--different, etc.). The principal trainer covered the list and gave it to a student not involved in testing, who then randomly ordered all the questions on the list. Each test included only one question on each topic. Second, test questions were presented intermittently during training sessions<sup>3</sup> on current (and thus unrelated) topics over several days until all questions on the test were presented. A numerical question thus may have occurred only once per week.<sup>4</sup> Only students who were working as trainers on novel topics acted as examiners. Questions were thus asked by students who had not trained Alex in the comprehension and conjunction studies (Pepperberg, 1990a, 1992), a system that further lessened the possibility of trainer-induced cuing (Pepperberg, 1981).

Multiple-topic tests not only avoid expectation cuing by the examiner and increase the general complexity of the task, but they prevent results from being affected by expectation cuing by the subject. That is, questions with a relatively restricted range of answers may allow a subject to perform better than is otherwise justified by its actual knowledge of the topic (see Discussion in Pepperberg, 1990c). My procedure ensured that Alex was never, in one session, tested exclusively on a single topic (e.g., numbers), on the same question (e.g., "What's same?") successively, or on questions with one particular correct response (e.g., color). A question was repeated in a session only if Alex's initial answer was incorrect (see Pepperberg, 1981, 1987b). Furthermore, the use of examiners who had not trained numerical labels lessened the possibility that Alex associated numerical test trials with particular humans (Pepperberg, 1987b).

*Numerical test procedure.* The specifics of the testing procedure resembled those in Pepperberg (1990a). During testing, four different quantities of items of two different colors and two different materials were placed on the same felt-covered tray that was used for studies on comprehension (Pepperberg, 1990a, 1992) and object permanence (Pepperberg & Kozak, 1986). To ensure further that the tray was not associated solely with the numerical task, we also placed various other collections of objects on the tray and intermittently asked questions as to colors, shapes, or materials of specific exemplars or collections of exemplars. We also used the same tray for training phoneme recognition.

A numerical trial began when the secondary trainer acting as examiner allowed Alex to touch his tongue to each of the exemplars that constituted the collection. This step enabled Alex to distinguish certain materials that are visually similar (e.g., Playdoh and plastic forms). The trainer then scattered these objects randomly, by heterogeneous handfuls, onto the surface of the tray. Only if an object was obscured by the placement of other objects or if the subsets obviously clumped together was the arrangement altered in any way. Each object was generally spaced less than 5 cm from the other nearest object.

The number of times an array was presented to Alex depended on his accuracy, which was determined as in previous studies (e.g., Pepperberg, 1981). When a secondary trainer questioned the bird, the principal trainer was present but sat in a corner of the room with her back to the bird. The principal trainer never asked the question or knew what was being presented. After Alex answered the secondary trainer, the principal trainer repeated what she heard the bird say. Her interpretation of Alex's response was unlikely to be influenced by hearing the type of question. In the previous conjunction study, for example, transcriptions of contextless tapes of Alex's responses in a session agreed with original evaluations to within 98.2%<sup>5</sup> (Pepperberg, 1992). If Alex produced the appropriate number label, he received praise and the objects to which the question referred or was allowed to request an alternative reward. There were then no further presentations of the same material during that test; 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 were test materials immediately, repeatedly presented, and presentation continued until a correct identification was made; errors were recorded.<sup>6</sup>

*Scoring.* Alex's test scores were calculated in two ways. Because the test procedure required that a question be repeated 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 overall test score (results 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.

To learn if Alex's results were statistically significant, we used two different values of chance. First, chance was based on the number labels relevant to the task (1 of 4), as if Alex was randomly guessing, without targeting a subset, among the quantities of all presented subsets. Second, chance was calculated not from a completely random choice of subsets but on the basis of one of the targeted categories of subsets (1 of 3), as if Alex were, for example, guessing the answer to "How many green truck?" after targeting only green items. In that case, he could respond to the number of (a) all green objects, (b) green trucks, or (c) the other green subset. All of these calculations are conservative in that they assume 100% comprehension of the labels that identify the subset. A less conservative calculation may include the probability of his misunderstanding a question and correctly identifying the quantity of a wrong subset. Given that his accuracy was about 80% on somewhat more complicated comprehension tasks (Pepperberg, 1990a, 1992), the probability for such a misidentification of subset was small but not nonexistent. Chance could also be based on the possible number labels (1 of 6), as if Alex was randomly guessing after limiting his choice to numbers (1-6) after hearing "How many ... ". Note, too, that Alex could produce any other of the numerous vocalizations in his repertoire; all calculations assume that Alex will always ( $p = 1$ ) attend and respond correctly to the "How many" part of the question.

## **Results**

### *Overall Results*

Calculated with respect to first-trial performance, Alex's overall accuracy was 45 of 54, or 83.3%. For all values of chance,  $p < .0001$  on binomial tests. Overall (first trials and correction trials), Alex's score was 54 of 65, or 83.1 %. Table 1 shows the data separately for each of the different numbers tested.

### *Individual Results*

Scores varied from 9 of 9, or 100%, on "four" to 6 of 9, or 66.6%, on "three." Table 1 presents the various significance values for binomial tests for calculations based on different values of chance; in all cases,  $p <$

.0341. How scores vary with respect to each number is particularly important. If Alex used a perceptual, rather than a counting, strategy, the number of errors would be significantly lower in the subitizing range (see Mandler & Shebo, 1982). Alex's errors (Table 1), however, varied randomly with respect to the number of items to be identified.

Because the pattern of such data cannot be tested for significance by a normal linear regression, I performed a regression instead on the severity of the error. According to Gallistel and Gelman (1992), the mechanism generally known as subitizing can be defined by data that conform at least qualitatively to Weber's law: The greater the reference numerosity, the more imprecisely the subject will distinguish between it and nearby numerosities. Thus a test (linear regression) to see if the severity of Alex's errors (distance between a correct answer and his response) increased as the number of items increased is not only a measure of his accuracy but also provides evidence for subitizing or counting (or for using preverbal or verbal mechanisms; Gallistel & Gelman, 1992). A sequential canonical analysis (Gorsuch & Figueredo, 1991), in which the direction of the error was partialled out (because the direction of error had to go toward the smaller numbers as the quantities increased,  $r = -.33$ ,  $F[1, 96] = 6.37$ ,  $p = .01$ ), showed that the severity of Alex's errors was not significantly correlated with the increase in the number of objects to identify. No correlations occurred in either the linear regression,  $F(1, 96) = 0.55$ , *ns*, or in a test for nonlinearity,  $F(4, 96) = 0.58$ , *ns*.

## Discussion

### *Nonnumerical Sources of Error*

Detailed examination of Alex's responses suggests that some errors may have been related to his processing of questions rather than his ability to quantify objects. As in other comprehension tasks (e.g., Pepperberg, 1992), errors could arise from four sources that were independent of the targeted competence. These sources were: (a) confusion of labels that sound alike; (b) misunderstanding a label that directs the search; (c) problems with respect to perceptual boundaries (e.g., differences in avian and human color perception); or (d) an inability to comprehend the conjunctive condition (i.e., to understand that information from two categories must be used to determine the object of the search).

Because almost half of the errors in an earlier recursion study (Pepperberg, 1990a) involved either comprehension or production of labels that, at least to humans, sound similar (e.g., "rock" and "block"), I attempted to eliminate such confounds in this task. A few trials did, however, combine block and truck. Although such labels are similar only in the final consonant cluster, Alex did on two occasions give the number of trucks, rather than of blocks. On one of these occasions, however, he erred in responding on the basis of the wrong color as well.

Errors from the second nonnumerical source cannot be distinguished from numerical errors. In previous studies, however, Alex's scores on tests that required comprehending label attributes were about 80% (Pepperberg, 1990a, 1992), and his scores in this study were comparable. On some questions he may have misinterpreted the defining labels and then correctly labeled the number of an incorrect subset. Indeed, 8 of 9 erroneous responses provided the correct number for an incorrect subset.

Of the 9 errors few are attributable to perceptual rather than cognitive deficits. In previous comprehension studies (Pepperberg, 1990a, 1992), collections could be combinations of such objects as rocks (Playdoh) and rawhide, which are difficult even for humans to distinguish by sight. No such combinations were used here. Also, psittacine and human visual systems differ in their color boundaries (see Bowmaker, 1986, personal communication, January 1993), and previous studies (Pepperberg, 1990b; 1992) have suggested that errors on orange are likely to be "rose" or "yellow" and errors on purple to be "blue" or



"rose." Such combinations were, for the most part, avoided in this study. In the two trials for which Alex responded with the number of the subset identified by the wrong color, he provided, in one, the number of green (rather than yellow) wool and, in the other, the number of yellow (rather than purple) papers.

**Table 1. Results and Errors for Numerical Trials**

Number	Score (of 9)	%	Probability		Erroneous labels used	Attribute
			$\frac{1}{4}$	$\frac{1}{3}$		
1	7	77.8	.0012	.0073	2, 3	C, O
2	8	88.9	.0001	.0009	4	R
3	6	66.7	.0087	.0341	2, 2, 4	O, O, O
4	9	100.0	<.0001	.0001		
5	7	77.8	.0012	.0073	2, 3	C-O, O
6	8	88.9	.0001	.0009	4	C

*Note.* Probabilities are based on chance values of  $\frac{1}{4}$  and  $\frac{1}{3}$ ; all probabilities can also be multiplied by a factor to take into account the probability that Alex was able to identify correctly the targeted subset. Attribute is the type of error: C = color (Alex's response was consistent with the number of items in the subset of the nontargeted color but targeted objects); O = object (Alex's response was consistent with the number of items in the subset of the nontargeted objects but targeted color); C-O = color and object (Alex's response was consistent with the number of items in the subset of the nontargeted color and objects); and R = random (in the sense that Alex's response did not relate to any presented numerical quantity).

In all cases Alex appeared to attend to the conjunctive condition. He never provided the number for a subset defined by only one attribute (e.g., the number of all the trucks). Only one error was random, that is, did not correspond to any presented quantity. In 8 of 9 errors, he responded with the quantity of a nontargeted subset that was defined by a conjunction of color and object labels.

Using sequential canonical analysis (Gorsuch & Figueredo, 1991), I specifically tested whether the number of Alex's errors on the attributes of color and object identity were correlated with his numerical errors and whether the number of color and object errors increased as the number of objects to label increased. The test found no significant correlation.

#### *Latencies to Respond*

Many studies use latencies to respond to distinguish between subitizing and counting. The rationale is that the mechanisms for these two processes require different time frames, specifically, that subitizing requires less time than counting (e.g., Mandler & Shebo, 1982). Latency to response is not, however, a viable criterion for Alex. As reported previously (Pepperberg, 1987a, 1987b, 1988a), his readiness to respond to a given set of exemplars is correlated to his level of interest in obtaining these items rather than any other factor. The familiarity of the exemplars in the numerical tasks generally precluded Alex's interest in obtaining the items; he generally requested a nut or a cork after completion of a trial and did not interact with the targeted exemplars. Latency of response therefore could be neither meaningfully measured nor compared with those of other subjects.

### **General Discussion**

#### *Numerical Competence*

At least three possible interpretations exist for the data. First, Alex may have been subitizing because his capacity for perceptual recognition of quantity greatly exceeds that of humans. Second, he may have

been counting according to the conditions described by Trick and Pylyshyn (1989) and Gallistel and Gelman (1992). Third, he may have been using some combination of these processes. Arguments can be made for each of these interpretations.

*Subitizing.* Alex may have been subitizing if avian numerical perceptual capacities are superior to those of humans in the visual mode. For example, Alex may have used perceptual capacities to segregate the set to be enumerated, which he then subitized (H. Davis, personal communication, Fall 1992; Dehaene & Changeux, 1993). Because the targeted items are scattered among 9 or more distractors, such a strategy seems unlikely (at least to humans), but the existence of such extended perceptual capacities cannot be dismissed. Avian numerical perception may surpass that of humans in the auditory, sequential mode (see Thompson, 1968, 1969; Wolframm & Todt, 1982); conceivably, such an ability may transfer not only between auditory and visual modes but also between sequential and simultaneous processes (e.g., Seibt, 1982). To ensure that Alex was beyond the subitizing range, we would have to test him on simultaneous visual quantities larger than those perceived sequentially in nature. Data on how gray parrots process natural vocalizations are, however, nonexistent.

*Counting.* If Alex does not have an unusually robust perceptual mode—that is, if his capacities are equivalent to those of humans—his data are consistent with counting rather than subitizing behavior. Gallistel and Gelman (1992), in their interpretation of subitizing, suggested that the property of the mental magnitude that represents numerosity (the distribution of magnitudes to which a digit maps) scales linearly with the mean magnitude of the numerosity. Thus, if Alex were subitizing, his accuracy for smaller numbers would be greater than for larger numbers, and a strong correlation would be expected between the severity of his errors and the increase in the actual quantities to be identified. However, his accuracy did not vary significantly, and no significant correlation was found between numerosity and the severity of errors. Moreover, according to Trick and Pylyshyn (1989), human subjects cannot use subitizing in a task, like that given Alex, that involves labeling the quantity of a subset of items distinguished from other subsets by a conjunction of qualifiers. Comparable performance cannot, however, ensure comparable mechanisms across species, because procedures that control for alternative explanations of behavior in one species may not provide controls in others.

*A combination of processes.* Because Alex was not under time constraints, he may have combined subitizing and counting processes. He may have used a perceptual process to overcome the distractor effect in order to subitize quantities less than or equal to 3 and then have switched to a more accurate counting strategy for 4 or more. Note the small rise in number (but not in severity) of errors at 3. Such a combination of strategies is particularly intriguing: Alex would then have shown a numerical competence not unlike that of humans at the higher quantities.

### *Effect of Language Training*

Gallistel and Gelman (1992), who separated what are generally considered subitizing and counting into preverbal and verbal processes, suggested that the ability to label quantities (through a process of mapping count words to preverbal magnitudes) distinguishes the two processes. Hurford (1987) and Dehaene (1992) have also suggested a close correlation between labeling and number skills, in the sense that numerical cognition is "a layered modular architecture, the preverbal representation of approximate numerical magnitudes supporting the progressive emergence of language-dependent abilities such as verbal counting" (Dehaene, 1992, p. 35). Did Alex's extensive training on labeling tasks enable him to exhibit behavior (e.g., Premack, 1983) that could not exist in a preverbal bird?

I have previously discussed this issue (Pepperberg, 1987b, 1990a; Pepperberg & Brezinsky, 1991; Pepperberg & Funk, 1990). Of utmost importance is the fact that although Alex uses, in appropriate

situations, abstract vocalizations at a representational level, his behavior is not, in most respects, isomorphic with human language. Clearly, his ability to label quantities provides a means to test his capacities in a manner similar to that used with humans, and his capacity to label may help him form a more precise representation of a quantity in a subset. Whether his ability to label carries him to Gallistel and Gelman's (1992) verbal level of counting is unclear. Rice (1980), for example, believed that linguistic input cannot teach a nonlinguistic concept for which a subject is unready. Thus, although Alex can use "six," for example, to label an appropriate number of objects, the label may still represent a perceived attribute of the subset (like color) rather than be the result of a counting process. According to Trick and Pylyshyn (1989), however, the presence of distractors prevents the use of such a perceptual strategy, at least in humans. Most likely, Alex's ability to separate out, visually collect, and enumerate items in the targeted subset is a consequence of the years of task-oriented training he has received; his ability to produce a vocal label for the result merely provided an extremely efficient testing mechanism.

In sum, the data demonstrate that Alex can vocally label the quantity of a specific subset of objects in a complex heterogeneous array. The subset is uniquely defined by the conjunction of one color and one object category and the array is designed to provide maximal confounds, or distractors, as in comparable work with humans (Trick & Pylyshyn, 1989). The data suggest that a nonhuman, nonprimate, nonmammalian subject has, at least on this task, an overt level of numerical competence equivalent to that of humans.

Whatever strategies Alex uses in this situation, he cannot technically count unless he understands that number labels represent an ordered set of tags that can be paired with each object. Studies to demonstrate explicitly Alex's sense of ordinality are underway because simple ordered vocal production (i.e., "1, 2, 3 ...") may indicate only rote memorization (e.g., Dehaene, 1992). A clear demonstration of 1:1 pairing is not simple to produce: Parrot claws and beaks are less suitable than primate digits for pointing (Pepperberg, 1988b), and training Alex to produce specific actions, such as small head movements in the direction of the targeted objects, may increase the likelihood of his use of rhythmic patterns or sequential subitizing (Davis & Perusse, 1988).

Determining whether Alex can count in the human sense, however, may not be especially useful. The ability can always be redefined to exclude whatever animals still cannot do (see Davis & Perusse, 1988; Fouts, 1978; Pepperberg, 1988c, 1990a). The value of such work lies instead in its use as a step toward determining the extent of animal cognitive competence, that is, whether the observed abilities can be developed into more complex capacities, such as labeling larger quantities and understanding addition and subtraction. In such a manner numerical competence may indeed become a comparative indicator of general intelligence.

## Notes

<sup>1</sup> Questions often arise about the replicability of a study that uses a single subject. If a single subject in a study reliably demonstrates a capacity, the implication is that this aptitude is within the capacity of the species (Pack, Herman, & Roitblat, 1991; Pepperberg & Funk, 1990; Triana & Pasnak, 1981; Wright, Cook, Rivera, Sands, & Delius, 1988). Of course, negative results for a single subject cannot prove the lack of aptitude for the species as a whole.

<sup>2</sup> Testing was interrupted between August 1990 and February 1991 because Alex became seriously ill (aspergillosis) in September 1990 and the lab was moved from Illinois to Arizona in late 1990. Testing was interrupted between May 1991 and May 1992 because of two bouts of psittacosis.

<sup>3</sup> Training sessions had to be in progress at least 5 min before a test question could be administered. Such a procedure ensured both that Alex was attending to his trainers and that the sessions would not be interrupted before several training trials had been completed. After this initial 5-min period, test questions could be given at any time.

<sup>4</sup> Questions sometimes arise as to whether the principal trainer was truly blind, that is, whether she was incapable of remembering the answer to specific questions on a given test. Note that to remember the correct answer, the principal trainer would have had to store in memory, usually over the course of 5-14 days, answers to each of 12 questions not only of the type "How many blue keys?" but "What is the shape of the green key?" (Pepperberg, 1992), "What object is blue?" (Pepperberg, 1990a), and "What color is bigger?" (Pepperberg & Brezinsky, 1991), as well as the answers to test questions that were simultaneously being asked of two juvenile parrots and to remember such answers for a course of 54 different sets of test arrays. This memory feat would also have had to occur in the midst of fast-paced training sessions on sets of completely different topics in which notes were constantly being made of correct and incorrect responses. Such considerations do not include other demands on memory load during these 5-14 days.

<sup>5</sup> This percentage represented 106 matches of 108 vocalizations. As a control, the principal trainer made 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 of a tape from which all questions had been edited. The two transcriptions of the student's vocalizations matched to within 95.8% (68 of 71 vocalizations).

<sup>6</sup> Occasionally the examiner rather than Alex erred. In about 1 in 20 trials (particularly during student exam periods), an examiner erred and scolded Alex for a correct response. Alex then repeated his correct response, despite the procedures, which encourage a lose-shift strategy. The examiner recognized her error, and the bird got his reward. Note that although this is not a formal blind test, it produces the same results.

## References

- Aoki, T. (1977). On the counting process of patterned dots. *Tohoku Psychologica Folia*, 36, 15-22.
- Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: A new look at visual numerosity judgments. *Perception*, 5, 327-334.
- Beckmann, H. (1924). Die Entwicklung der Zahlleistung bei 2-6 jährigen Kindern [The development of numerical ability in 2- to 6-year-old children]. *Zeitschrift für Angewandte Psychologie*, 22, 1-72.
- Bowmaker, J. K. (1986, June). *Avian color vision and the environment*. Paper presented at 19th International Ornithological Congress, Ottawa, Ontario, Canada.
- Boysen, S. T., & Berntson, G. G. (1988, April). *Four is more: Evidence for an understanding of ordinality in a chimpanzee*. Paper presented at the 58th Annual Meeting of the Midwestern Psychological Association, Chicago.
- 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). New York: Cambridge University Press.
- Boysen, S. T., & Capaldi, E. J. (1992). *The development of numerical competence: Animal and human models*. Hillsdale, NJ: Erlbaum.

- Bums, R. A. (1988). Subitizing and rhythm in serial numerical investigations with animals. *Behavioral and Brain Sciences, 11*, 581-582.
- Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that comprise it. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 3-17.
- Case, R. (1985). *Intellectual development: Birth to adulthood*. San Diego: Academic Press.
- Cole, M., & Scribner, S. (1974). *Culture and thought: A psychological introduction*. New York: Wiley.
- Davis, H. (1984). Determination of the number three by a raccoon (*Procyon lotor*). *Animal Learning & Behavior, 12*, 409-413.
- Davis, H., & Albert, M. (1986). Numerical discrimination by rats using sequential auditory stimuli. *Animal Learning & Behavior, 14*, 57-59.
- Davis, H., & Albert, M. (1987). Failure to transfer or train a numerical discrimination using sequential visual stimuli in rats. *Bulletin of the Psychonomic Society, 25*, 472-474.
- Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence, and a new research agenda. *Behavioral and Brain Sciences, 11*, 561-615.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1-42.
- Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*, 399-407.
- Fouts, R. (1978). Sign language in chimpanzees: Implications of the visual mode and the comparative approach. In F. C. C. Peng (Ed.), *Sign language and language acquisition in man and ape* (pp. 121-136). Boulder, CO: Westview Press.
- Fuson, K. C., & Hall, J. W. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 49-107). San Diego: Academic Press.
- Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In W. E. Kessen, A. Ortony, & F. I. M. Craik (Eds.), *Thoughts, memories, and emotions: Essays in honor of George Mandler* (pp. 65-81). Hillsdale, NJ: Erlbaum.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43-74.
- Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number* (2nd ed.). Cambridge, MA: Harvard University Press.
- Glanville, A. D., & Dallenbach, K. M. (1929). The range of attention. *American Journal of Psychology, 41*, 207-236.
- Gorsuch, R. L., & Figueredo, A. J. (1991, October). *Sequential canonical analysis as an exploratory form of path analysis*. Paper presented at the American Evaluation Association conference, Chicago.
- Giittinger, H. R. (1979). The integration of learnt and genetically programmed behaviour: A study of hierarchical organization in songs of canaries, green finches, and their hybrids. *Zeitschrift für Tierpsychologie, 49*, 285-303.
- Hurford, J. R. (1987). *Language and number*. Oxford, United Kingdom: Basil Blackwell.
- Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3*, 281-282.
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.
- Lenneberg, E. H. (1971). Of language, knowledge, apes and brains. *Journal of Psycholinguistic Research, 1*, 1-29.
- Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*, 1-22.
- Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature, 315*, 57-59.
- Meek, 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.

- Miller, G. (1956). The magical number  $7 \pm 2$ : Some limits on our capacity to store information. *Psychological Bulletin*, 63, 81-97.
- Oyama, T., Kikuchi, T., & Ichihara, S. (1981). Span of attention, backward masking and reaction time. *Perception and Psychophysics*, 29, 106-112.
- Pack, A. A., Herman, L. M., & Roitblat, H. L. (1991). Generalization of visual matching and delayed matching by a California sea lion (*Zalophus californianus*). *Animal Learning & Behavior*, 19, 37--48.
- Pepperberg, I. M. (1981). Functional vocalizations by an African grey parrot (*Psittacus erithacus*). *Zeitschrift für Tierpsychologie*, 55, 139-160.
- Pepperberg, I. M. (1983). Cognition in the African Grey parrot: Preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning & Behavior*, 11, 179-185.
- Pepperberg, I. M. (1987a). Acquisition of the same/different concept by an African grey parrot (*Psittacus erithacus*): Learning with respect to color, shape, and material. *Animal Learning & Behavior*, 15, 423--432.
- Pepperberg, I. M. (1987b). Evidence for conceptual quantitative abilities in the African Grey parrot: Labeling of cardinal sets. *Ethology*, 75, 37-61.
- Pepperberg, I. M. (1987c). Interspecies communication: A tool for assessing conceptual abilities in the African grey parrot. In G. Greenberg & E. Tobach (Eds.), *Cognition, language and consciousness: Integrative levels* (pp. 31-56). Hillsdale, NJ: Erlbaum.
- Pepperberg, I. M. (1988a). Comprehension of "absence" by an African grey parrot: Learning with respect to questions of same/different. *Journal of the Experimental Analysis of Behavior*, 50, 553-564.
- Pepperberg, I. M. (1988b). An interactive modeling technique for acquisition of communication skills: Separation of "labeling" and "requesting" in a psittacine subject. *Applied Psycholinguistics*, 9, 59-76.
- Pepperberg, I. M. (1988c). Studying numerical competence: A trip through linguistic wonderland? *Behavioral and Brain Sciences*, 11, 595-596.
- Pepperberg, I. M. (1990a). Cognition in an African gray parrot (*Psittacus erithacus*): Further evidence for comprehension of categories and labels. *Journal of Comparative Psychology*, 104, 41-52.
- Pepperberg, I. M. (1990b). Referential mapping: A technique for attaching functional significance to the innovative utterances of an African Grey parrot (*Psittacus erithacus*). *Applied Psycholinguistics*, 11, 23--44.
- Pepperberg, I. M. (1990c). Some cognitive capacities of an African grey parrot. In P. J. B. Slater, J. S. Rosenblat, & C. Beer (Eds.), *Advances in the study of behavior* (Vol. 19, pp. 357--409). San Diego: Academic Press.
- Pepperberg, I. M. (1992). Proficient performance of a conjunctive, recursive task by an African gray parrot (*Psittacus erithacus*). *Journal of Comparative Psychology*, 106, 295-305.
- Pepperberg, I. M., Brese, K. J., & Harris, B. J. (1991). Solitary sound play during acquisition of English vocalizations by an African grey parrot (*Psittacus erithacus*): Possible parallels with children's monologue speech. *Applied Psycholinguistics*, 12, 151-178.
- Pepperberg, I. M., & Brezinsky, M. (1991). Acquisition of a relative class concept by an African gray parrot (*Psittacus erithacus*): Discriminations based on relative size. *Journal of Comparative Psychology*, 105, 286-294.
- Pepperberg, I. M., & Funk, M. S. (1990). Object permanence in four species of psittacine birds: An African gray parrot (*Psittacus erithacus*), an Illiger minimacaw (*Ara maracana*), a parakeet (*Melopsittacus undulatus*), and a cockatiel (*Nymphicus hollandicus*). *Animal Learning & Behavior*, 18, 97-108.
- Pepperberg, I. M., & Kozak, F. A. (1986). Object permanence in the African grey parrot (*Psittacus erithacus*). *Animal Learning & Behavior*, 14, 322-330.
- Premack, D. (1976). *Intelligence in ape and man*. Hillsdale, NJ: Erlbaum.
- Premack, D. (1983). The codes of man and beast. *Behavioral and Brain Sciences*, 6, 125-176.

- Rice, M. (1980). *Cognition to language: Categories, word meanings, and training*. Baltimore: University Park Press.
- Seibt, U. (1982). Zahlbegriff und Zählverhalten bei Tieren: Neue Versuche und Deutungen [Number concepts and number behavior by animals: New experiments and interpretations]. *Zeitschrift für Tierpsychologie*, *60*, 325-341.
- Smith, W. J. (1988). Patterned daytime singing of the eastern wood-pewee, *Contopus virens*. *Animal Behaviour*, *36*, 1111-1123.
- Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*, 97-127.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49). New York: Wiley.
- Taves, E. H. (1941). Two mechanisms for the perception of visual numerosness. *Archives of Psychology*, *37*, 1-47.
- Thompson, N. S. (1968). Counting and communication in crows. *Communications in Behavioral Biology*, *2*, 223-225.
- Thompson, N. S. (1969). Individual identification and temporal patterning in the cawing of common crows. *Communications in Behavioral Biology*, *4*, 29-33.
- Trick, L., & Pylyshyn, Z. (1989). *Subitizing and the FNST spatial index model* (Tech. Rep. No. COGMEM 44). London, Ontario, Canada: University of Ontario.
- Trick, L., & Pylyshyn, Z. (1991). *A theory of enumeration that grows out of a general theory of vision: Subitizing, counting, and FINSTs* (Tech. Rep. No. COGMEM 57). London, Ontario, Canada: University of Ontario.
- Triana, E., & Pasnak, R. (1981). Object permanence in cats and dogs. *Animal Learning & Behavior*; *9*, 135-139.
- von Glasersfeld, E. (1982). Subitizing: The role of figural patterns in the development of numerical concepts. *Archives de Psychologie*, *50*, 191-218.
- Wolfgang, J., & Todt, D. (1982). Pattern and time specificity in vocal responses of blackbirds *Turdus merula* L. *Behaviour*; *81*, 264-286.
- Wolters, G., van Kempen, H., & Wijnhuizen, G. (1987). Quantification of small numbers of dots: Subitizing or pattern recognition? *American Journal of Psychology*, *100*, 225-237.
- Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior*; *16*, 436-444.

## Appendix

### Sample of Trials for Each Number

Trial	Exemplars	Question	Response
1	1 orange chalk, 2 orange wood, 4 purple wood, 5 purple chalk	How many purple wood?	4
2	1 yellow block, 2 gray block, 4 yellow wool, 6 gray wool	How many yellow block?	1
3	1 rose wood, 2 blue nail, 3 blue wood, 5 rose nail	How many rose nail?	5
4	2 gray truck, 3 gray key, 4 orange key, 5 orange truck	How many gray key?	2, 3
5	1 blue box, 3 green box, 4 blue cup, 6 green cup	How many green cup?	6
6	1 purple rock, 2 green rock, 3 purple plastic key, 4 green plastic key	How many green rock?	2

*Note.* For these particular trials, *wood* refers to tongue depressors, *wool* to woolen pompons, *rock* to Playdoh tubes, *nail* to nails partially covered with plastic mollies. Singular labels were used for all questions to avoid cuing *one*.