PART I

THE COMPARATIVE APPROACH

Prior to Darwin, humans lived in a different world from other species. While our machines were inhabited by ghosts, other creatures were simply machines devoid of internal states (Descartes, 1641). With the publication of the *Origin of Species* in 1859, however, people began to question this anthropocentric assumption of a discontinuity between "us" and "them." Thirteen years later in Darwin's final book, *The Expression of the Emotions in Man and Animals* (1872), he developed this argument of continuity between human and nonhuman species further by drawing on observations of parallels of expression and reaction in a wide range of species. It is fair to say that *The Expression of the Emotions* led directly to the development of "comparative psychology" and provided legitimacy to the study of animal behavior as a means to better understand ourselves (Workman, 2013). In 1894, Conway Lloyd Morgan formalized this approach in his book, *Introduction to Comparative Psychology*, setting out the ground rules for the comparative method.

During the twentieth century, comparative psychology subdivided into two main approaches. One approach focused on the internal states of animals and eventually developed into the new field of animal cognition, while the other attempted to exorcise mentalistic language from the field of animal behavior and eventually became known as *behaviorism*. Today, studies of animal cognition still draw on Darwin's conception of continuity between species, but, by integrating developments in ethology and neuroscience, they also relate specific cognitive abilities to the behavioral ecology of a population.

It is fair to say that Irene M. Pepperberg's research into avian cognition is a major contribution to the field of animal cognition. Pepperberg's work with African Grey parrots showed how a species that has not shared a common ancestor with our own lineage since the late Carboniferous period can nonetheless exhibit human-like vocal communication. We begin Part I with her chapter on human–Grey parrot comparisons in cognitive performance.

The old adage that "elephants never forget" is based on a large body of anecdotal evidence. In recent years, however, field and lab studies have begun to put some flesh on these anecdotal bones. Lucy Bates has spent a number of years observing and testing this social giant. Her chapter on the cognitive abilities in elephants reinforces and dispels the myths that have built up around the intellectual prowess of these largest of all land animals.

Another group of animals that, due to their apparent complex social behavior, has long fascinated us is the cetaceans. Like elephants, whales and dolphins are renowned for their apparent cognitive prowess. In the third and final chapter in Part I, Ellen C. Garland and Luke Rendell consider culture and communication among cetaceans. Is it possible that we can improve our understanding of the roots of human language and culture by studying creatures that evolved in the three-dimensional liquid world of the oceans? Garland and Rendell provide us with the current state of play with regard to these questions and suggest fertile areas for future research.

REFERENCES

Darwin, C. (1859). *On The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: HarperCollins.

Descartes, R. (1641). *Meditations on First Philosophy*, trans. by J. Cottingham (1996). Cambridge, UK: Cambridge University Press.

Morgan, C. L. (1894). *Introduction to Comparative Psychology*. London: Walter Scott.

Workman, L. (2013). *Charles Darwin: Mindshaper*. Basingstoke: Palgrave Macmillan.

1   Human–Grey Parrot Comparisons in Cognitive Performance

IRENE M. PEPPERBERG

1.1 INTRODUCTION

Animal cognition is, by itself, an incredibly broad field, encompassing a huge variety of taxa and involving many different topics in both the laboratory and nature. When asked to view the field through the lens of an evolutionary perspective relative to humans, most scientists focus on our nearest relatives, the great apes (e.g., note the preponderance of chapters devoted to nonhuman primates in Vonk & Shackelford, 2012). Convergent evolution, however, provides striking insights into how distantly related species have responded to similar social and ecological challenges, and comparisons between avian and primate species have demonstrated remarkable parallels in various capacities (Emery & Clayton, 2004; Pepperberg, 1999, 2013). In areas such as vocal learning, avian species, separated from humans by over 300 million years of evolution (e.g., Hedges et al., 1996), actually provide a better model for study than do apes (e.g., Bolhuis & Everaert, 2013; Chakraborty et al., 2015). Such vocal ability can also lead to the acquisition of a limited form of referential symbolic communication with humans, which in turn can further facilitate comparative studies of avian–human cognitive capacities (e.g., Pepperberg, 1999). Interestingly, researchers have recently demonstrated that avian neural systems are functionally comparable to those of the great apes (e.g., Güntürkün & Bugnyar, 2016; Olkowicz et al., 2016), thereby providing a clear basis for advanced avian cognitive capacities. Grey parrots (*Psittacus erithacus*), which are often at the center of studies on avian intelligence and communication (Pepperberg, 1999), will be the focus of this chapter.

First, however, a bit of background: I have studied the cognitive and communicative abilities of Grey parrots for almost 40 years. Via a modeling procedure – the model/rival or M/R technique, adapted from experiments by both Bandura (1971) and Todt (1975) – I have trained these birds to acquire some level of referential communication using the sounds of English speech, then used this communication code to examine their cognitive abilities. My oldest subject, Alex, learned to use vocal labels to identify a wide variety of objects, colors, and shapes, processed queries to judge category, relative size, quantity, and the presence or absence of similarity/difference in attributes, and showed label comprehension and a zero-like concept; he demonstrated some understanding of phonological awareness and numerical competence more comparable to that of young children than to other nonhumans (Pepperberg, 2012b). His requests (for specific objects or to be moved to a particular location) were intentional (Pepperberg, 1988). My younger birds (particularly Griffin) are acquiring similar vocal abilities, particularly with respect to labeling of objects, colors, and shapes (e.g., Pepperberg & Nakayama, 2016; Pepperberg & Shive, 2001; Pepperberg & Wilcox, 2000).

This chapter reviews their recent advances. For Alex, the focus will be on his final numerical studies: inferring the cardinality of new number labels from their order on the number line (Pepperberg & Carey, 2012) and the addition of small quantities (Pepperberg, 2012a). His data provide evidence for actual counting (see Section 1.2). For Griffin, the focus will be on tests involving delayed gratification (the ability to forgo an immediate reward to gain one of better quality; Koepke, Gray, & Pepperberg, 2015), the recognition of optical illusions involving amodal and modal completion (respectively, recognizing occluded shapes and Kanizsa figures; Pepperberg & Nakayama, 2016), and (along with Grey parrots belonging to a colleague) tasks requiring reasoning by exclusion (inferring where an item is hidden after being given information on where it is not; Pepperberg et al., 2013, 2019). In most of these studies, Grey parrots demonstrate capacities comparable to those of children aged three to five years.

1.2 NUMBER CONCEPTS: COUNTING

Some understanding of number is a widespread phenomenon (in nonhumans, from fish [Petrazzini et al., 2015] to bears [e.g., Vonk & Beran, 2012]; in humans, even in preverbal children [Wynn, 1990] and preliterate hunter–gatherer societies [e.g., Frank et al., 2008]). However,

symbolic representation of number – the recognition that individual symbols represent exact, specific quantities – involves advanced capacities, once thought to be limited only to humans (reviewed in Pepperberg & Carey, 2012). Only a few nonhumans have demonstrated exact symbolic number representation: two apes, Matsuzawa's Ai (Matsuzawa, 1985) and Boysen's Sheba (Boysen & Berntson, 1989), and my subject, a Grey parrot, Alex (Pepperberg, 1987, 1994). Symbolic representation is important because it is a prerequisite for true counting, as defined by Gallistel and Gelman (1992) via several counting principles (CPs). CPs state that numerals must be applied in order to items in a set to be enumerated and in a one-to-one correspondence, that the last numeral in a count represents a set's cardinal value, and that the successor function (that each numeral is known to be exactly one more than the one before it and exactly one less than the one after it; e.g., Carey, 2009) must be understood. Determining whether nonhumans could acquire CPs has attracted widespread scientific interest.

Acquisition of CPs is not easy, and even children take several years to accomplish the task (Carey, 2009; Fuson, 1988). At about two years of age, most children can rattle off a series of number words (a "count list"), but they often confuse the order of the number labels and assign exact cardinal meaning only to "one"; other numerals mean "some" or "plural" (e.g., Barner & Bachrach, 2010). About nine months later, they learn "two"; other numerals are "more than two." A few months later, they master "three," then "four" (Wynn, 1992); they acquire a stable, accurate count list. Only then (generally somewhere between 3.5 and 4 years of age) do children *induce* the CPs and understand that each successive numeral in their count list is exactly one more than its predecessor. This induction separates them from those understanding only the exact meaning of a few small numbers ("one-," "two-," "three-," or "four-knowers"; Sarnecka & Carey, 2008); they can now encode cardinal value expressed by *any* numeral in their count list. Carey and I (Pepperberg & Carey, 2012) tested whether a Grey parrot could perform this induction; at the time, no nonhuman had been able to acquire this knowledge purely by induction.

We worked with Alex, who had previously been taught to use English count words ("one" through "sih" [six]) to label sets of one to six individual items exactly (production and comprehension; Pepperberg, 1987, 1994; Pepperberg & Gordon, 2005). He had also been taught to use the same count words to label Arabic numerals 1 through 6, but had had no training to associate any Arabic numeral with any specific quantity of items. Without training, he had subsequently inferred the relationship between the Arabic numerals and the sets of objects (Pepperberg, 2006b); that is, he had deduced the ordinality of the Arabic numerals by recognizing that an Arabic symbol had the same numerical value as its vocal label, comparing representations of quantity for which the labels stood and inferring their rank ordering based on these representations. So, for

example, given a green "5" and a yellow "2," he could answer both queries: "What color [is the] number [that is] bigger/smaller?" Notably, he had never been trained to recite number labels in order, nor had he even learned the number labels in order (see later in this section). Thus, even at this point, he appeared to exhibit numerical understanding far closer to that of children than other animals. However, he differed from humans and was like other nonhumans in that he had demonstrated no savings in his learning of larger numerals; that is, unlike four-year-old children, he did not acquire "five" and "six" via the successor function, without training. Why was Alex unlike children in this instance? Might the issue be Alex's difficulty not in learning the meaning of the numerical symbols, but rather in producing the English sounds? To generate any given English label, he had to learn to coordinate his syrinx, tracheal muscles, glottis, larynx, tongue height and protrusion, beak opening, and even esophagus (Patterson & Pepperberg, 1998). Might there be a way to dissociate vocal and conceptual learning to test this possibility?

Carey and I devised the following experiment (Pepperberg & Carey, 2012): I began by teaching Alex to identify vocally the Arabic numerals 7 and 8 in the absence of their respective quantities, divorcing the time needed to learn the speech patterns from any concept of number. After the labels were produced clearly, I trained him to understand that $6 < 7 < 8$; that is, where the new numerals fit on the number line. Without further training, he demonstrated that he understood the relationships among 7 and 8 and his other Arabic labels. Could he now, like children, spontaneously understand that "seven" represented one more physical object than "six," and that "eight" represented two more than "six" and one more than "seven," by labeling appropriate physical sets on first trials? Nothing in his training up to this point would provide specific information about the values of 7 and 8; they could refer to 10 and 20 items, respectively. Interestingly, all of his other numerals had been taught as either +1 or –1 than those he already knew (i.e., his first number labels were "three" and "four"; he was next taught "five" and "two," then "six" and "one"; Pepperberg, 1987, 1994); could he thus use past and present information to *induce* the cardinal meaning of the labels "seven" and "eight" – which he had learned to use for the Arabic numerals 7 and 8 – from their ordinal positions on an implicit count list? The answer was positive: Alex, like children, had created a representational structure that allowed him to encode the cardinal value expressed by any numeral in his count list (Carey, 2009); that is, to understand the successor function (NB: details of experimental design, including all controls for possible cuing, are in Pepperberg & Carey, 2012). Notably, the nonhuman primates tested so far (e.g., Boysen et al., 1993; Inoue & Matsuzawa, 2009) had to be trained to understand the ordinality of their numbers and have not yet inferred cardinality from ordinality.

This study provided the first demonstration that a nonhuman could engage in the bootstrapping process that underlies the construction of the integer list representation of number. Very much like young children, Alex could accomplish this task because he had true symbolic representation of his numerals, an ability fostered by his training in referential communication.

1.3 NUMBER CONCEPTS: ADDITION

Alex had also demonstrated a limited ability with respect to addition. By replicating part of a study on apes (Boysen & Berntson, 1989), I showed that he could watch his trainers hide two small sets of objects (quantities from zero to six) and, while they were still hidden, provide the *exact* vocal label for the total (i.e., requiring summation and symbolic representation of a hidden quantity; Pepperberg, 2006a). In contrast to most other addition studies with nonhumans, the task avoided use of only one token type of a standard size (e.g., whole marshmallows), which could allow evaluations based on contour and mass, not number (see discussion in Mix, Huttenlocher, & Levine, 2002). Overall, his data were comparable to those of young children (Mix et al., 2002) and, because he added to six, were more advanced than those of apes (Boysen & Hallberg, 2000).

Once Alex had acquired the numerals through 8, I returned to the addition task to learn if he could, like apes (Boysen & Berntson, 1989), sum three separate hidden sets or a set of two hidden Arabic numerals (Pepperberg, 2012a). These tasks would demonstrate further knowledge of the representational nature of the numerals. Addition of three sets would require two updates in memory rather than one; the study with Arabic numerals would determine whether he could spontaneously transfer to summing representations of quantities rather than physical quantities. As with the sets of items, he was sequentially shown two Arabic numerals initially hidden under cups and, in their consequent absence, was asked to vocally produce a label to indicate their sum. In a separate small set of trials, he was shown the same stimuli in the same manner, but was simultaneously presented with various Arabic numerals of different colors in randomized numerical order and asked for the color of the numeral representing the sum; the colors and positions of the numerals changed on each trial. The second set of trials ensured that Alex could not learn a particular pattern over time (e.g., "if I see X + Y, I say Z"). This procedure, with its additional step, would allow testing of the same sums many more times without training him to produce a specific response, unlike tasks given to other nonhuman subjects (Pepperberg, 2012a).

Because of Alex's death, the study did not contain enough trials to test all possible sums and combinations of addends or to repeat most queries. However, he received at least one trial for each sum from 1 to 8 for the Arabic numerals and at least one trial for each sum from one to six for the three sets of objects (just by chance, the trials requiring summation to eight were scheduled for the end of the study and could not be completed). His results were statistically significantly correct (Pepperberg, 2012a). The lack of replication of the various sums over the trials emphasizes the first-trial nature of the results and shows that no training could have been involved. Notably, if Alex's numerals had only approximate meanings, his errors would likely have exhibited a range close to the correct response. In contrast, such was the case only once (Pepperberg, 2012a). Overall, his data surpassed what would be expected if he were using the kinds of systems employed by most nonhumans or preverbal infants – for example, analog magnitude systems or object files, which cannot represent any positive integer above 4 exactly (for a review, see Carey, 2009).

Alex had only three trials on queries requiring a color response. The small number of trials preclude real statistical power, but tended toward significance: he erred on the first trial, but was correct on the next two. The task was unlike any he had ever experienced before and he had received no prior training; thus, he might not initially have understood the point of the task. In contrast to the ape that had previously been studied (Boysen & Berntson, 1989), however, Alex had to indicate the label not for the sum, but rather for the color of the numeral that represented the correct numerical sum (an additional step), and the total summed quantity on which he was tested could reach 8. His results thus suggested an intriguing level of competence on yet another numerical task – again, one based on his capacity for symbolic representation (once more, all experimental details can be found in the published paper – Pepperberg, 2012a).

1.4 DELAYED GRATIFICATION: THE "MARSHMALLOW TEST"

Delay of gratification involves postponing immediately available rewards to gain more desirable future rewards, maintaining the choice during delay, and tolerating the frustration of this self-inflicted delay. For example, children (about four years of age) are told they can eat one marshmallow placed in front of them immediately, but that they could earn a second if they refrain from eating the first until the experimenter returns from running an errand (Mischel, Shoda, & Rodriguez, 1989). As such, the task purportedly tests cognitive capacities related to future planning and self-control, capacities often thought to be lacking in nonhumans (Emery & Clayton, 2004). Delayed gratification has thus become an important tool for studying comparative cognition, particularly from an evolutionary perspective (Hillemann et al., 2014; Koepke et al., 2015). Interestingly, even many children have trouble passing the test, suggesting that it indeed examines important cognitive processes.

The task might appear straightforward, but actually involves multiple competing strategies. In nature, subjects

must evaluate whether the risks (e.g., losing out to competitors, experiencing energy depletion, succumbing to predation) outweigh the advantages of waiting to find a better or larger food source (Stephens, Kerr, & Fernández-Juricic, 2004). In the laboratory, subjects might fail because, based on personal experience, they do not trust the experimenter to fulfill the promise of the better or larger reward, for example (Kidd, Palmeri, & Aslin, 2013).

Delayed gratification also consists of two independent but related components, originally tested together in children (Mischel, 1974). The first, *delay choice*, is the initial election to wait for the better future reward. The second, *delay maintenance*, is the ability to bridge the delay interval; that is, to inhibit continuously the impulse to take the immediate reward during the delay after the initial choice to wait (Mischel et al., 1989; Toner & Smith, 1977; Toner, Lewis, & Gribble, 1979). Many studies on nonhumans test these components independently, such that in some experiments subjects first choose between an immediate, lower-value food or delayed, higher-value food and cannot change their decision during the experimenter-specified delay. Results may be confounded because subjects are often unable to inhibit initial pointing to the better payoff (Boysen & Berntson, 1995; Hillemann et al., 2014), and such tasks do not test whether subjects can sustain their delay choice. Other experiments, using either an exchange or accumulation paradigm, do allow subjects to alter their initial choice at any time. In the former experiments, which involve waiting for *better*, subjects are given a less preferred food that they can keep intact throughout the delay and then exchange for something of greater value or end the trial by consumption. In the latter experiments, which involve waiting for *more*, a series of identical edibles is moved within the subject's reach at a fixed rate (usually seconds) per item. Food accumulates until the subject interrupts by taking the available rewards (for a review, see Koepke et al., 2015).

Interestingly, results may differ depending on whether the task involves waiting for *better* or *more*. Nonhuman primates succeed at both tasks, often waiting for up to 10 minutes; birds, however, generally succeed only for *better*, rarely waiting longer than a few seconds for *more* (for reviews, see Hillemann et al., 2013; Koepke et al., 2015). The reasons for these species differences are unclear, but they may involve differences in foraging behavior (Koepke et al., 2015). Grey parrots, like other previously tested avian species, had shown little aptitude for waiting for *more* (Vick, Bovet, & Anderson, 2010), but had never been tested on *better*. My students and I thus decided to examine this possibility.

We were particularly interested to see how Griffin would perform because, unlike previously tested nonhumans, he already had a rudimentary understanding of the vocal label "wait" (Koepke et al., 2015). Every day at noon he is given highly preferred cooked grains, initially too hot to eat, and is told to "wait," although hearing the label does not decrease his anticipatory actions. He evinces similar actions at my entry into the lab, expecting to be picked up and preened; he thus hears "wait" while I use hand sanitizer and remove my outdoor shoes. Again, the instruction has little effect on his behavior, but he may have associated the vocalization with a delay of something he desires. Thus, unlike most other nonhumans, avian or primate (e.g., Auersperg, Laumer, & Bugnyar, 2013; Dufour et al., 2007, 2012), he did not have to be trained extensively on the exchange paradigm usually used to test *better* before beginning the experiment; that is, learning to exchange a nonfood token or less preferred food item for something desirable – activities that may have affected the responses of other subjects (e.g., showing effects of training). Here, the question was whether Griffin would be able to *infer* that "wait" could be associated with an *alternative* choice.

Our study differed from others given to nonhumans in several additional ways (Koepke et al., 2015). First, unlike previously tested birds (Auersperg et al., 2013; Dufour et al., 2012; Hillemann et al., 2014), Griffin was not subjected to longer delays as the experiment progressed; he had no idea how long a given trial might take as trials *randomly* lasted 10, 40, 160, 320, 640, or 900 seconds. Thus, he was not trained to wait for progressively longer periods, but had to choose to wait as long as necessary. Second, given that reward visibility made waiting more difficult for children (Mischel & Ebbensen, 1970; Mischel et al., 1989), we tested Griffin in both visible and nonvisible conditions for each time delay. For both cockatoos (Auersperg et al., 2013) and corvids (Dufour et al., 2012; Hillemann et al., 2014), the more preferred reward was always visible unless cached; caching corvids were more successful (Dufour et al., 2012). Success with a nonvisible reward would suggest the need to maintain a mental representation of the preferred food during delays. Third, we introduced control trials for each time delay in which we presented the favored item first and asked Griffin to wait for a less favored one. If he understood that "wait" was an option and not a command, he should appropriately ignore our instruction and fail to wait on these trials. Success would show that he was not simply trained to wait until he received a cue from the experimenter. Finally, some experimenters (e.g., Bramlett et al., 2012; Drapier et al., 2005) have used much less desirable items as the immediate reward. The subjects in these studies might easily have lost interest in the items, the temptation to consume them might have been missing, and the point of delaying gratification would be less relevant. Here, we used seven items that Griffin strongly preferred over his basic diet, but for which he still had a range of preference; these items were carefully ranked and we never contrasted the most and least favorable items. Using a range of rewards also prevented Griffin from considering a particular treat as a signal to wait, forcing him to evaluate his choice on each and every trial, including control trials. He had to make a new decision for each set of items, further demonstrating that he understood the metrics of the task. As

a separate issue, we also ensured that Griffin's ability to wait was not a consequence of satiation by using small rewards (e.g., half or less of a cashew) and performing only a few trials, separated by at least 15 minutes, on any given day. More than four trials per day occurred only when Griffin gave evidence of desiring more rewards (e.g., repeatedly vocally requesting a high-tier treat).

Griffin was successful, waiting for the preferred reward for up to 15 minutes (Koepke et al., 2015). For all delays on visible and nonvisible conditions, he waited on approximately 90 percent of the trials, and his success rate did not depend on delay length or reward pairing, and neither did it vary significantly with trial number (i.e., he was not learning the task). On all of the control trials – one in each of the different delay times and conditions – Griffin elected to eat the initial, preferred reward rather than wait for the less preferred item, therefore demonstrating his attention to reward type and his recognition of "wait" as a label for an action rather than a command. Furthermore, he did not wait because he had lost interest in the reward: on two 900-second (15-minute) delays, he failed once after waiting 740 seconds and once after waiting 815 seconds. He seemed to recognize the concept of delayed gratification, even at the onset of the experiment.

Of particular interest were his coping strategies for dealing with the delays, many of which were reminiscent of those used by children (Mischel & Ebbesen, 1970). He could not cover his eyes with his claws the way children would cover their eyes with their hands, but he might throw the cup containing the less favored treat across the room so as not to have it in view or, like children, move the cup just out of reach or talk to himself. He might preen extensively, and, like children, try to fall asleep while waiting. He also, like children (Steelandt et al., 2012), occasionally licked the treat, but did not eat it. When observed in children, these behavior patterns are often termed "self-distractive." Unlike corvids, which can cache (Dufour et al., 2012), hiding the immediate reward was not part of Griffin's repertoire. Thus, at least in a quality exchange paradigm, little difference exists among various avian species and primates (NB: experimental details are in Koepke et al., 2015). We still, however, must test whether Griffin can wait for *more* to claim total equivalence across species on delayed gratification overall.

Although other species, as noted above, have indeed succeeded on this task without experience in symbolic representation, our ability to test Griffin in exactly the same way as children – a procedure unlike that used for most other nonhumans – was dependent upon his comprehension of the verbal symbol "wait."

1.5 ROBUST RECOGNITION OF SHAPES: MODAL AND AMODAL COMPLETION

One of the few clear markers of higher-order cognition is the ability to transfer concepts across domains (Rozin, 1976). Alex demonstrated such behavior with respect to

absence and a zero-like concept (Pepperberg & Brezinksy, 1991; Pepperberg & Gordon, 2005), but only to a limited extent with concepts such as shape. Although he could transfer from uncolored three-dimensional (3D) wooden polygons of a single size to 3D polygons of various materials, sizes, and colors, he was limited to 3D regular polygons. Could Griffin, similarly trained, now demonstrate invariance over a much wider range of testing conditions, transferring, for example, from 3D objects to 2D drawings? Furthermore, could he transfer to the phenomenon of occluded objects (amodal completion) and imaginary, subjective contours (Kanizsa figures) as shown in Figure 1.1? Such testing images differ in more numerous and varied ways from training exemplars than those given to Alex, and for subjective contours, no such overlap occurs.

This study (Pepperberg & Nakayama, 2016) was particularly interesting because demonstrating such abilities in nonhumans has been difficult. Although neither a primate brain nor a visual system would seem a prerequisite for such capacities – results from many studies, from insects to nonhuman primates, are consistent with various animals responding appropriately to 2D objects that are visible to humans as partially occluded or as partly represented in their outline form by subjective contours – these studies are subject to a variety of alternative interpretations (reviewed in Pepperberg & Nakayama, 2016). However, with a parrot that understood symbolic representation – in other words, could vocally identify various shapes ("#-corner," where # = 1, 2, 3, 4, 6) – we could recreate a situation more like that given to humans, who in most cases are given a small number of trials involving several different stimuli and simply asked to label what they see. Furthermore, Griffin grew up in a very rich environment, more similar to that of preschool children. Thus, he saw and manipulated real-world 3D objects of all forms, materials, and colors, both in full view and occluded, during a period of over 16 years before the study commenced. Such experience may be a prerequisite for carrying out the tasks being studied (Stephan, Wilkinson, & Huber, 2013).

We asked Griffin to label paper depictions of stimuli, something he had never before been asked to do. To test for amodal completion (occlusion), we used variously
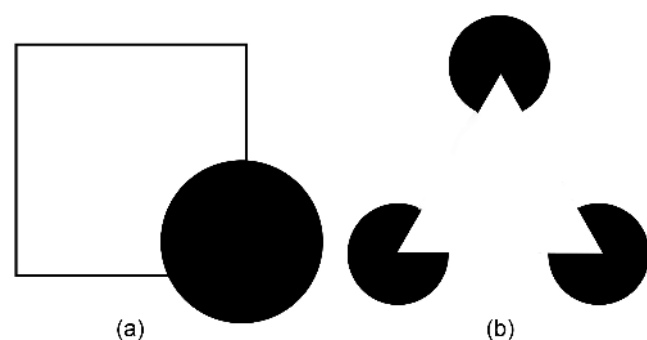

Figure 1.1 (a) Occluded shapes (amodal completion) and (b) subjective (Kanizsa, illusory) shapes (modal completion).
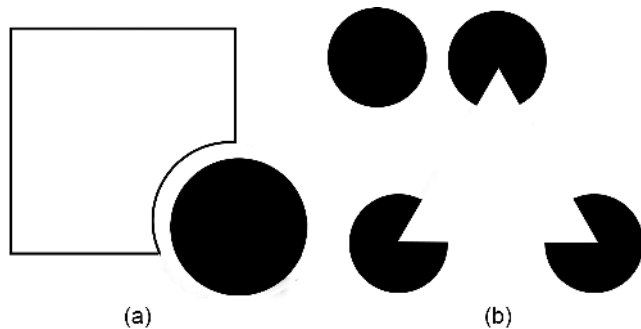
Figure 1.2 (a) Detached probe for amodal completion and (b) probe for modal completion.

colored regular polygons for each shape he could label, of different sizes, occluded by black circles (e.g., asking "What shape blue?"). Griffin had not been trained to label either circles or the color black, making it unlikely that he would attempt to label them. We also used other black shapes as occluders, to see if being able to label occluders would distract him from the task at hand. To ensure that Griffin was responding appropriately to occluded figures, we also asked him to identify irregular polygons ("detached probes") that were not occluded (e.g., regular polygons that looked as though a bite had been removed from them, with an appropriately sized adjacent circle; Figure 1.2a). For the subjective contours, Kanizsa figures were constructed using black "pac-men" to form regular polygons, again for each shape he could label, of different sizes, on colored paper. Controls ("probes," one or two for each of the #-cornered shapes) involved placing additional circles or "pac-men" near the Kanizsa figure so that Griffin could not simply quantify black objects (e.g., Figure 1.2b). Each trial was unique with respect to color, size of polygon, or size of occluder/pac-men. He was given only 38 trials for each type of task.

We decided that the strongest way to test Griffin's abilities would be to track his responses to single presentations of each possible probe. Most experiments in the nonhuman literature use repeated presentation of identical probes for testing (e.g., Nagasaka & Wasserman, 2008; Nakamura et al., 2011) and either reward the subject for all probe trials (potentially encouraging guessing) or, after decreasing primary rewards to a set percentage similar to the proportion of probe trials, reward none of the probe/test trials (potentially discouraging possible correct attempts). Our procedure would avoid any possible issues of familiarity, training, or encouraging either guessing or discouraging correct attempts.

Griffin identified the figures for amodal and modal completion at statistically significant levels and was correct on all probe trials. Importantly, he was correct on his very first trials, showing that no training was occurring and that transfer from 3D to 2D figures was immediate. Interestingly, for the detached probes (e.g., Figure 1.2a), where no occlusion occurred, he responded to figures never before seen (irregular polygons) with the number

of visible corners. For modal test stimuli, in contrast to amodally completed stimuli, there is *nothing* in common at an image level between the trained and test stimuli, yet his accuracy was identical to that for the amodal stimuli. He also was not quantifying pac-men or numbers of circles, because in no instance did an error correspond to the quantity of these objects (NB: experimental details can be found in Pepperberg & Nakayama, 2016).

Against Griffin's remarkable success, a substantial literature exists showing that, as noted above, with very few exceptions, animals either do not show these completion phenomena at all or show some degree of success only after having undergone considerable training with very closely related stimuli (reviewed in Pepperberg & Nakayama, 2016). In some instances where training was not an issue, success or failure may have involved mechanisms unrelated to the perceptual ones being examined here, and performance may have actually relied on mechanisms that do not match those of humans (e.g., luminance, aspect ratio: Minini & Jeffery, 2006; responses based on numbers of stimuli, stimulus generalization, reliance on local cues such as one angle in the stimuli, statistical averaging over thousands of trials: reviewed in Pepperberg & Nakayama, 2016). In other instances, actual tasks differed considerably among the laboratories (with respect to, e.g., motion, 2D vs. 3D stimuli, CRT vs. LCD monitors [i.e., flicker fusion effects] and pre-exposure to stimuli) with the consequence that results also varied considerably (again, reviewed in Pepperberg & Nakayama, 2016).

What could account for Griffin's success? Many other creatures must be able to solve problems involving at least some form of amodal completion in their daily lives (e.g., Lea, Slater, & Ryan, 1996; Regolin & Vallortigara, 1995). For example, processing partial clues about a potential predator and reacting is safer than not, even if some false alarms incur costs. Modal completion may rely on similar, early-level visual processing (e.g., imagine three black circles on a colored background that are occluded by a triangle of the same color; Nakayama, He, & Shimojo, 1995). As noted earlier, many species have demonstrated abilities consistent with, although not conclusively defined by, such processing. Griffin, in contrast, performed in a manner more sophisticated than other laboratory-based subjects. His results may be a consequence of two capacities, although others may also be involved. First, he understood symbolic representation: that a sound could stand for a physical object. Thus, his understanding that a 2D depiction could stand for a 3D representation of reality, including depth perception, would not be surprising. Note that baboons failed when tested for amodal completion in a task involving considerable training on 2D figures and forced, two-choice testing on 2D transfer stimuli (Deruelle et al., 2000), but did succeed (although only after several hundred training trials and 60 forced, two-choice testing trials) when given both training and testing stimuli that provided background depth cues indicating

that the occluder was indeed in front of the targeted object (Fagot et al., 2006; see also Nagasaka, Brooks, & Wasserman, 2010, on bonobos). Second, as noted earlier, Griffin was raised in an extremely rich environment for a laboratory subject, providing him with the same kind of experience that seems to enable young children to succeed on such tasks (Kellman & Spelke, 1983).

1.6 STUDIES ON EXCLUSION

Inference by exclusion implies the ability to base a decision on the exclusion of potential alternatives. The task appears simple, but actually involves several levels of competence. The simplest level is based purely on avoidance. For example, a subject with experience in picking A when told "take *A*" and then shown A and B and told "take *B*" may avoid A merely because it simply knows not to choose A in the absence of hearing the appropriate cue, "*A*"; such a level of exclusion does not, however, necessarily lead to any understanding of the inferential relationship between "*B*" and B (Dixon, 1977). Likewise, a task may involve hiding two items in two containers, with one container subsequently being shown to be empty; to show strong inferential reasoning by exclusion, the subject's choice of the full container must not simply be a consequence of avoiding the container from which something has been extracted or the one most recently manipulated (reviewed in Mikolasch, Kotrschal, & Schloegl, 2012; Pepperberg et al., 2013). Even young children (approximately three years old) succeed on these relatively simple versions of the task, but their level of understanding is unclear (Hill, Collier-Baker, & Suddendorf, 2012). In more complicated tasks (e.g., Premack & Premack, 1994), the subject is shown that two *different* items are hidden, one is removed surreptitiously and shown to the subject, and the subject must then infer where the other must be. In the strongest sense, the subject also must be shown not to be working under the "maybe A, maybe B" assumption described by Mody and Carey (2016): in the two-position hiding, the subject might assume that either position is possible and eliminate one based on evidence, but not *conclusively* understand that B is the *only* logical choice, and rather assume that now it is just the more *likely* one.

My students and I tested four Grey parrots on tasks designed to examine at least some of their abilities regarding this form of reasoning. We note that previous studies with this species showed that the task is exquisitely sensitive to the conditions under which it is run. For example, very few of the birds in the Mikolasch, Kotrschal, and Schloegl (2011) study (which replicated that of Premack & Premack, 1994) succeeded, but their birds had to distinguish between identical cups and thus use spatial memory to infer where the remaining object was hidden – an additional cognitive process. In another experiment, auditory cues were used (either the empty or full cup was shaken), and Grey parrots

succeeded only when the container was shaken horizontally, not vertically (Schloegl et al., 2012), probably because the vertical motion resembled that of the feeding behavior of a parent and thus provided a distracting confound.

In an attempt to clarify the abilities of Grey parrots on this task, my students and I performed two experiments on four birds (for additional details about the parrots, see Pepperberg et al., 2013). Two of the birds, Griffin and Arthur, had lived in a laboratory for almost all of their lives. Griffin was 16 years old at the time of the experiment and had demonstrated full object permanence (OP; a prerequisite for studies on exclusion) as a juvenile (Pepperberg, Willner, & Gravitz, 1997); he had been in the lab since he was 7.5 weeks old. Arthur was 13 years old and had been in the lab since he was 1 year old. Arthur had not been formally tested on OP, but, like Griffin, he had been the subject of a previous study that ensured he understood that the vocal phrase "Go pick up cup" was a request to make a choice and that items that had been hidden under the cups should still be available unless contrary evidence existed; he also had had training on referential speech. Two other Grey parrots, Pepper and Franco, lived in a suburban household (that of the Hartsfields) with two adult humans who had previously been trainers in the Pepperberg lab. Pepper, a female, was 15 years old and had lived with the Hartsfields since she was about 3 months old; Franco, a male, was 10 years old and had joined the household when he was 7 years old, having lived with another family previously. Pepper had received considerable training on referential communication, but had not been formally tested on production or comprehension; Franco had entered the Hartsfield household with the capacity to produce some human speech, but his referential knowledge was unknown. He subsequently had about three years of referential training, but no formal testing. Thus, we could examine whether extensive laboratory experience was a prerequisite for succeeding on the task. None of the birds were food nor water deprived.

The first experiment mostly replicated the Grey parrot study of Mikolasch, Kotrschal, and Schloegl (2011), which (as noted above) had replicated that of Premack and Premack (1994), to learn if our subjects could succeed on the basic task. Here, parrots watched an experimenter hide two equally desirable foods under two separate opaque cups (but unlike the previous study, using cups of two different colors), surreptitiously remove one of the foods, and then, in view of the birds, pocket/eat that food, leaving the birds to find the still baited cup. The experiment contained controls for various alternative explanations for the birds' behavior (e.g., olfactory cues, local enhancement, human cues, simple association rules based on what treat is handled; Mikolasch et al., 2011; Pepperberg et al., 2013). All birds succeeded at statistically valid levels (Pepperberg et al., 2013), but might they still have avoided

a cup from which something had been removed rather than tracking the eaten/handled food?

Specifically, might the birds conceivably have focused on the lack of removal of *an* object, not on *the* specific object? Would the birds still be willing to return to a cup from which a favored reward was partially removed in contrast to a cup that held something to which they were indifferent? Thus, in a second experiment, some trials were run with one food slightly more preferred than the other, during which *two* items of each type were hidden in each cup and only one of the items was removed from one cup – randomly, either the more or less preferred food. If the birds understood what was happening, they should always go to the cup with the preferred food, as even one preferred food item would be a better result than two of the less preferred items. Sessions also included Experiment 1-type trials to see if the birds tracked when to use and when not to use exclusion. Thus, the birds would be rewarded for attending closely to all of the experimental aspects needed to infer how to receive their preferred treat.

All four birds succeeded on these trials, demonstrating that they did not simply avoid a cup from which something had been removed (Pepperberg et al., 2013). Three birds (Griffin, Arthur, and Pepper) also appeared to be able to switch between experimental conditions; that is, to understand something about when to use or when not to use exclusion. The fourth bird (Franco) also succeeded on these trials starting with two hidden items in each cup, but seemed to have difficulty switching between these trials and Experiment 1-type trials, actually making several errors on the latter. Possibly he was not closely attending and expected two treats to be in each cup on each trial, or possibly he lost interest in the task and ceased to attend closely to the procedure (for a discussion of "boredom" in Grey parrots, see Pepperberg & Gordon, 2005).

The results of these experiments demonstrate that Grey parrots can reason by exclusion to some degree. They do not simply avoid an empty container, but infer the most advantageous choice based on the specific context of the trial; they are at least at the level of three-year-old children. The data also showed that laboratory experience did not seem to be critical for success, although it is possible that at least some training in symbolic representation assisted all four birds. Hill, Collier-Baker, and Suddendorf (2012) have specifically argued for a relationship between capacities for symbolic representation and success on the cup task used here. The rationale is that subjects having undergone symbolic training could transfer their deductive/inferential reasoning across tasks, representing the hidden object(s) in some manner. (Again, all experimental details can be found in the published paper – Pepperberg et al., 2013.) Recently, my colleagues and I have shown that Grey parrots could pass the stronger test of exclusion, as defined and executed by Mody and Carey (2016), which involves four hiding places (two pairs) and probabilistic reasoning. Here, a reward is hidden in one cup of each pair (e.g., A, C), and one cup (e.g., B) is then shown to be empty. The

subject should conclude that the reward is 100 percent likely in A and only 50 percent likely in either C or D, and so choose A. To ensure that Griffin was not simply choosing the cup next to the empty one, we performed an additional experiment in which for some trials he was given an incentive to gamble (to choose C or D) by being shown that a special, much preferred treat was being hidden in the 50-percent side. Griffin succeeded at all tasks (for details, see Pepperberg et al., 2019). His success places him beyond the level of five-year-old children.

1.7 CONCLUSION

All of these studies have demonstrated that Grey parrots can perform advanced cognitive tasks, comparable to those performed by nonhuman primates and young children. In Alex's number studies and Griffin's experiments on modal and amodal completion, success required their capacities for symbolic representation. Studies on delayed gratification and exclusion also could be seen to involve some aspects of such representation – for the former task, knowledge of the meaning of "wait"; for both tasks, mental representations of treats that were not visually present. Pepper and Franco, although not given the same formal training as Alex and Griffin, also had experience in labeling, as did Arthur (note, however, that Arthur's success in such labeling tasks was limited; Pepperberg & Wilkes, 2004). Symbolic training, which can be independent of vocal learning, has enabled several species to succeed on advanced cognitive tasks that would not likely otherwise have been possible (e.g., apes: Boysen, 2006; monkeys: Livingstone et al., 2014; Grey parrots: Pepperberg, 1999, 2013; Pepperberg & Carey, 2016). Premack (1983, 1984) specifically claimed that such training enhances nonhuman primates' abilities on specific tasks such as those requiring abstract judgment and analogic reasoning. Granted, many studies examining advanced cognitive capacities in nonhumans do not rely on training in symbolic representation (for reviews, see ten Cate & Healy, 2017; Vonk & Shackelford, 2012; Wasserman & Zentall, 2006), but such training does allow nonhumans to be tested in ways that are very similar to those used with humans, and thus provides a particularly strong method for examining comparative cognition via an evolutionary perspective.

REFERENCES

Auersperg, A. M. I., Laumer, I. B., & Bugnyar, T. (2013). Goffin cockatoos wait for qualitative and quantitative gains but prefer "better" to "more". *The Royal Society: Biology Letters*, **9**, 20121092.

Bandura, A. (1971). Analysis of modeling processes. In A. Bandura, ed., *Psychological Modeling*. Chicago, IL: Aldine-Atherton, pp. 1–62.

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Science*, **60**, 40–62.