Volume 93, February 2013

ISSN 0376-6357 BPRODA 93 1–166 (2013)



Behavioural Processes

SCIENCE



Special issue:

Functional Relationships: A Tribute to the Contributions of Anthony A. Wright Guest Editors: Jeffrey S. Katz and Jonathon D. Crystal

BEHAVIOURAL PROCESSES (2013) Volume 93

CONTENTS					
Special Issue Functional Relationships: A Tribute to the Contributions of Anthony A. Wright					
Guest Editors: Jeffrey S. Katz and Jonathon D. Crystal					
Editorial J.S. Katz and J.D. Crystal (USA) Functional relationships	1				
Special Issue Papers A.A. Wright (USA) Functional relationships for investigating cognitive processes	4				
L.C. Elmore, A.D. Passaro and A.A. Wright (USA) Change detection for the study of object and location memory	25				
J.F. Magnotti, A.M. Goodman, T.A. Daniel, L.C. Elmore, A.A. Wright and J.S. Katz (USA) Visual object complexity limits pigeon short-term memory					
J.D. Crystal (USA) Remembering the past and planning for the future in rats	39				
B.M. Basile and R.R. Hampton (USA) Monkeys show recognition without priming in a classification task	50				
D.M. Kelly and J.F. Reichert (Canada) Angular amplitude matters: Exploring the functional relationship of geometric cue use by Clark's nutcrackers (Nucifraga columbiana)	62				
K.D. Bodily, Z.A. Kilday, C.K. Eastman, K.A. Gaskin, A.A. Graves, J.E. Roberts and B.R. Sturz (USA) Beacons and surface features differentially influence human reliance on global and local geometric cues when reorienting in a virtual environment	71				
<i>I.M. Pepperberg (USA)</i> Abstract concepts: Data from a Grey parrot	82				
D.J. Merritt and E.M. Brannon (USA) Nothing to it: Precursors to a zero concept in preschoolers	91				
R.G. Cook, A.A. Wright, and E.E. Drachman (USA) Categorization of birds, mammals and chimeras by pigeons	98				
J.J. Nieworth (USA) Chasing sounds	111				
<i>M. Jitsumori, N. Nakamura and E.A. Wasserman (Japan, USA)</i> Discrimination of coherent and incoherent motion by pigeons: An investigation using a same-different motion discrimination task	116				
L. Castro and E. A. Wasserman (USA) Humans deploy diverse strategies in learning same-different discrimination tasks	125				
K.J. Leising, J.E. Wolf and C.M. Ruprecht (USA) Visual discrimination learning with an iPad-equipped apparatus	140				
<i>R.P. Kesner (USA)</i> Role of the hippocampus in mediating interference as measured by pattern separation processes	148				
J. Bachevalier, A.A. Wright and J.S. Katz (USA) Serial position functions following selective hippocampal lesions in monkeys: Effects of delays and interference	155				

Contents lists available at SciVerse ScienceDirect



Behavioural Processes



journal homepage: www.elsevier.com/locate/behavproc

Editorial Functional relationships

Identifying functional relationships between variables is fundamental to experimental psychology. In order to discover functional relationships, a critical (independent) variable must be identified, manipulated along a continuum, and its effect on another (dependent) variable analyzed. In comparative cognition, such parametric variations are rare because of the added investment necessary once some critical variable is identified. However, when the investment is made the impact can be enormous. The discovery of functional relationships has been the foundation of Anthony A. Wright's research program.

The Comparative Cognition Society annually recognizes at the International Conference on Comparative Cognition a scientist who has had a major impact on the field of animal cognition with the Comparative Cognition Society Research Award. On March 9, 2012 Tony Wright was the honoree, with a special symposium in honor of his contributions to comparative cognition. The symposium presentations represent in part the impact Wright has had on his students and colleagues. Each presenter (Bachevalier, Cook, Crystal, Katz, and Kesner) contributed an article to this Special Issue honoring Tony. In addition a collection of invited manuscripts further show his impact.

Wright's featured article in the Special Issue documents many of his academic greatest hits. To this we add that Tony grew up South of Los Angeles spending time along the coast of California (including Catalina Island) where he developed an appreciation for surfing, fast cars, and sailing. He made his way up the coast and received his BA from Stanford University in 1965. At Stanford he was influenced by a number of eminent scientists, and it was a young Gordon Bower that inspired Tony to devote his graduate studies to experimental psychology. Tony and his soon-to-be wife Susan moved from Stanford California (The Farm) to New York City (The Big Apple) where Tony received his PhD from Columbia University in 1971. At the time, Columbia was a hot bed for operant conditioning and Tony developed his passion for parametric manipulation to reveal functional relationships. Tony became an assistant professor at the University of Texas at Austin, but within a year he and Susan headed to Houston where Tony rose through the ranks at the University of Texas Health Science Center at Houston (which includes the University of Texas Medical School at Houston), where he remains today. In 1983, Tony spent 8 months on sabbatical leave in New Zealand on a Fogarty Senior International Fellowship. In 1989, Tony spent 12 months on sabbatical leave in Germany as a Recipient of the Alexander von Humboldt Prize. He is a Fellow of the prestigious Society of Experimental Psychologists.

Wright's research career has several distinct intertwined paths all bound together by a systematic understanding of functional

relationships in comparative cognition. Wright's dissertation, 'Psychometric and psychophysical hue discrimination functions for the pigeon' along with his 'Color naming functions for the pigeon' were his first forays into revealing functional relationships. These studies (along with hue-discrimination studies in rhesus monkeys) focused on color discrimination and represent the first major area of his career. Sections of his dissertation were published in a number of premier journals at the time: Journal of the Experimental Analysis of Behavior, Psychological Review, and Vision Research (Wright, 1972a, 1972b, 1972c, 1974). Another key hallmark of Wright's career is his love for designing and building the apparatus. Wright considers the best way to study some cognitive/perceptual issues (e.g., color perception) is to design an ideal environment and build the specialized apparatus, as opposed to what is more typically done by starting with an apparatus and figuring out what experiment can be conducted with this commercial testing equipment. The blue print can be seen on the cover of this Special Issue and right below the blue print is the optical apparatus for the color-naming experiment. This optical apparatus made its way from New York to Austin and then to Houston and is affectionately known as the "Taj" (Bill Cumming, Tony's Columbia mentor at the time, named it the "Taj Mahal" of Skinner boxes), due to the detail and beauty of the craftsmanship.

The second path of Wright's career involves memory processing. He and his colleagues developed the first visual and auditory list memory procedures for non-human primates (Sands and Wright, 1980a, 1980b, 1982; Wright and Rivera, 1997). The visual list memory experiments used slide projectors (cutting-edge technology at the time) and were automatically controlled by a 64-kbyte CPM Cromemco computer (the left side of the Special Issue cover shows a Cromemco computer controlling tape decks for monkey auditory list memory studies). The visual list-memory functions collected in Wright's laboratory have focused on capuchin monkeys, rhesus monkeys, pigeons, and humans (Wright et al., 1985; Wright, 1999). Wright and his collaborators collected a vast amount of travel slides to conduct the research which have become an instrumental set of stimuli in his research. Not to be lost in today's image-bountiful internet, back in the day when members of the Wright laboratory traveled, their 35-mm cameras were by their sides looking for new stimuli from around the world to be used in their experiments. The research Wright and his collaborators conducted has stood the test of time (like the color perception functions) and the body of work is considered among the most important research for proactive and retroactive interference memory processes in nonhumans and humans. In order to compare pigeons and monkeys to humans, Wright needed nonverbal stimuli. Humans verbally label travel slides, and they can use verbal processes to rehearse and



Fig. 1. A social network of Anthony Wright's published works. Line width between Wright and authors identified in each node is proportional to the number of papers coauthored with Wright. The darker the shade of green of a node indicates more publishing with Wright. The social network was constructed by John Magnotti.

recognize the images. To solve the verbal coding problem and level the 'playing field' Wright developed the use of kaleidoscope patterns for human memory studies and photographed all of the images (>1100) himself. The body of this memory research spans many of the top journals including *Journal of Experimental Psychology: Animal Behavior Processes, Journal of Experimental Psychology: General, Psychological Science*, and the crowning jewel is the cover of *Science* (seen on the cover of the Special Issue). To this continuing body of work, Wright's latest endeavor into memory processes was to develop a change detection procedure for nonhumans to explore visual short-term memory in monkeys and pigeons (Wright et al., 2010; Elmore et al., 2011).

The third path of Tony's research involves abstract-concept learning in nonhumans. Here Tony and colleagues designed experiments that showed nonhumans can learn abstract concepts if the conditions are suitable (Wright, 1997; Wright et al., 1988; Wright and Katz, 2006). Such systematic variation was critical to revealing abstract-concept learning, as many researchers had concluded

that many animal species could not learn abstract concepts based on "failures to find". Wright (1997) showed via manipulating the sample observing response (FR) in a matching-to-sample procedure that pigeons would show increasing transfer to novel stimuli with increases in the sample FR. With a small FR pigeons showed no concept learning (i.e., chance performance on novel images) but with a large FR (20) they showed full concept learning (i.e., transfer accuracy to novel images was equivalent to baseline accuracy). To conduct this research, Wright designed and built a pigeon operant chamber whereby the stimuli were presented from the floor of the chamber instead of from a front stimulus panel. This way pigeons pecked down at the stimuli and also ate grain placed on top of the computer images. The chamber can be seen on the cover of the Special Issue along with "cartoon" stimuli used in these experiments. Using computer generated cartoon stimuli was another important advance in animal studies by Wright. Researchers had traditionally used small stimulus sets because the commercial in-line projectors had only 12 stimuli (e.g., color circles, line tilts). Using computer

graphics software to construct stimuli allows one to produce a virtual unlimited set of stimuli (see the duck, apple, and grape on the Special Issue cover). Sure enough, Tony was able to show that another parameter critical in revealing abstract-concept learning is the training set size. Here Tony and his colleagues have shown that by increasing the training set, full concept learning will emerge in pigeons as well as capuchin monkeys and rhesus monkeys (Wright and Katz, 2006).

Fig. 1 illustrates a social network of Wright's collaborations created by John Magnotti. The social network was generated by inputting the coauthors from all of Wright's peer reviewed papers and book chapters. Each node represents a coauthor. The darker the shade of green of a node indicates more articles published with Tony. The width of the lines is proportional to the number of papers coauthored between each researcher in the network. The network also reveals different collaborative eras to a degree with Sands and Cook to the right and Katz to the lower left of the Tony Wright node.

The articles in this Special Issue further represent Wright's impact on comparative cognition. The authors consist of former graduate students, postdoctoral fellows, and other colleagues influenced by Wright. The articles demonstrate the breadth of impact Wright has had ranging from visual working memory (Elmore, Passaro, and Wright; Magnotti, Goodman, Daniel, Elmore, Wright, and Katz), memory systems (Crystal; Basile and Hampton; Kelly and Reichert; Bodily, Kilday, Eastman, Gaskin, Graves, Roberts, and Sturz), numerical concepts (Pepperberg; Merritt and Brannon), categorization (Cook, Wright, and Drachman; Neiworth; Jitsumori, Nakamura and Wasserman; Castro and Wasserman), discrimination learning (Leising, Wolf, and Ruprecht) and the neural processes of memory (Kesner; Bachevalier, Wright, and Katz). The featured article by Wright (2013) describes his ongoing research program that has exploited the study of functional relationships to develop new insights into comparative cognition.

We would like to thank all the contributors to the Special Issue for allowing us to put together a terrific collection of articles. We appreciate the assistance of many of the authors that helped review articles. Lastly, we would like to thank the Comparative Cognition Society for allowing us to serve as guest editors of the Special Issue and honoring Tony.

References

- Elmore, L.C., Ma, W.J., Magnotti, J.F., Leising, K.K., Passaro, A.D., Katz, J.S., Wright, A.A., 2011. Visual short-term memory compared in rhesus monkeys and humans. Curr. Biol. 21, 975–979.
- Sands, S.F., Wright, A.A., 1980a. Primate memory: retention of serial list items by a rhesus monkey. Science 209, 938–940.
- Sands, S.F., Wright, A.A., 1980b. Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. J. Exp. Psychol. Anim. Behav. Process. 6, 386–396.
- Sands, S.F., Wright, A.A., 1982. Monkey and human pictorial memory scanning. Science 216, 1333–1334.
- Wright, A.A., 1972a. Psychometric and psychophysical hue discrimination functions for the pigeon. Vision Res. 12, 1447–1464.
- Wright, A.A., 1972b. The influence of ultraviolet radiation on the pigeon's color discrimination. J. Exp. Anal. Behav. 17, 325–337.
- Wright, A.A., 1972c. Construct a monochromator from a single interference filter. J. Exp. Anal. Behav. 18, 61–63.
- Wright, A.A., 1974. Psychometric and psychophysical theory within a framework of response bias. Psychol. Rev. 81, 322–347.
- Wright, A.A., 1997. Concept learning and learning strategies. Psychol. Sci. 8, 119–123. Wright, A.A., 1999. Visual list memory in capuchin monkeys (*Cebus apella*). J. Comp.
- Psychol. 113, 74–80. Wright, A.A., 2013. Functional relationships for investigating cognitive processes. Behav. Process., http://dx.doi.org/10.1016/j.beproc.2012.11.003.
- 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. Anim. Learn. Behav. 16, 436–444.
- Wright, A.A., Katz, J.S., 2006. Mechanisms of same/different concept learning in primates and avians. Behav. Process. 72, 234–254.
- Wright, A.A., Katz, J.S., Magnotti, J.F., Elmore, L.C., Babb, S., Alwin, S., 2010. Testing pigeon memory in a change detection task. Psychon. B Rev. 17 (2), 243–249, http://dx.doi.org/10.3758/PBR.17.2.243.
- Wright, A.A., Rivera, J.J., 1997. Memory of auditory lists by rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 23, 441–449.
- Wright, A.A., Santiago, H.C., Sands, S.F., Kendrick, D.F., Cook, R.G., 1985. Memory processing of serial lists by pigeons, monkeys, and people. Science 229, 287–289.

Jeffrey S. Katz* Auburn University, United States

Jonathon D. Crystal Indiana University, United States

* Corresponding author. E-mail address: KATZJEF@auburn.edu (J.S. Katz) Contents lists available at SciVerse ScienceDirect





Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Functional relationships for investigating cognitive processes

Anthony A. Wright*

University of Texas Health Science Center, Medical School at Houston, United States

ARTICLE INFO

Article history: Received 18 July 2012 Received in revised form 8 November 2012 Accepted 9 November 2012

Keywords: Comparative cognition Psychophysics Learning Memory Pigeons Monkeys Humans

ABSTRACT

Functional relationships (from systematic manipulation of critical variables) are advocated for revealing fundamental processes of (comparative) cognition-through examples from my work in psychophysics, learning, and memory. Functional relationships for pigeon wavelength (hue) discrimination revealed best discrimination at the spectral points of hue transition for pigeons-a correspondence (i.e., functional relationship) similar to that for humans. Functional relationships for learning revealed: Item-specific or relational learning in matching to sample as a function of the pigeons' sample-response requirement, and same/different abstract-concept learning as a function of the training set size for rhesus monkeys, capuchin monkeys, and pigeons. Functional relationships for visual memory revealed serial position functions (a 1st order functional relationship) that changed systematically with retention delay (a 2nd order relationship) for pigeons, capuchin monkeys, rhesus monkeys, and humans. Functional relationships for rhesus-monkey auditory memory also revealed systematic changes in serial position functions with delay, but these changes were opposite to those for visual memory. Functional relationships for proactive interference revealed interference that varied as a function of a ratio of delay times. Functional relationships for change detection memory revealed (qualitative) similarities and (quantitative) differences in human and monkey visual short-term memory as a function of the number of memory items. It is concluded that these findings were made possible by varying critical variables over a substantial portion of the manipulable range to generate functions and derive relationships.

© 2012 Elsevier B.V. All rights reserved.

This article was inspired by suggestions from the nominating committee for a research-award presentation (2012 Comparative Cognition Society meeting). These suggestions included: Organize your presentation (and this article) around a theme that has guided your work (hence the title), include earlier work that some (younger) members of the audience may be unfamiliar, and mention individuals that have influenced your work. Therefore, I begin by mentioning individuals that shaped my interests and thinking, followed by: A psychophysical study of pigeon color vision; two concept learning studies with pigeons and monkeys; visual list memory studies with monkeys; a proactive-interference memory study with pigeons; and a visual short-term memory study with monkeys and humans in change detection.

As an undergraduate at Stanford University, I was greatly influenced by Gordon H. Bower (see Fig. 1). I became interested in experimental psychology and comparative cognition from an introductory psychology course, taught by Richard C. Atkinson with

E-mail address: anthony.a.wright@uth.tmc.edu

the introductory text book written by Stanford professor Ernest R. Hilgard. Hilgard and Atkinson were deeply involved in learning and emphasized animal learning and comparative studies of learning. My laboratory introduction was an animal learning course taught by Gordon Bower who recently had arrived from Yale University where he conducted his graduate work with Neal Miller (see Bower, 2011). Bower's learning course was a revelation for me. I and the other students in the course conducted a series of experiments with our own rats and were introduced to learning issues, many of which are still actively pursued today. Following the learning course, I conducted a research project with Professor Bower related to his influential learning model (Bower, 1962), involving different item types, conditions, and interference (a topic revisited later in this article). When it came time for me to apply for graduate studies, Gordon advised me to go to Columbia University which had strong programs in experimental psychology.

I arrived at Columbia University in 1965. The psychology department was located in Schermerhorn Hall (since its inception in 1898, and still is, see Fig. 2). The inscription above the doorway reads "FOR THE ADVANCEMENT OF NATURAL SCIENCE 'SPEAK TO THE EARTH AND IT SHALL TEACH THEE''. Research at Columbia was concentrated in two areas of experimental psychology (legacies of James McKeen Catell, the first Chair of Psychology, and Robert Sessions Woodworth, Catell's student). One area was visual psychophysics—headed by

^{*} Correspondence address: Department of Neurobiology and Anatomy, University of Texas Houston Medical School, 6431 Fannin Street, Suite 7.174, Houston, TX 77030, United States. Tel.: +1 713 500 5627; fax: +1 713 500 0623.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.003

Gordon H. Bower Stanford University (1960-1965)

Fig. 1. Gordon Bower: undergraduate mentor at Stanford University.

Professor Clarence H. Graham. The other area was animal learning which included William (Bill) W. Cumming (with whom I began my studies), William N. (Nat) Schoenfeld, and Herbert (Herb) S. Terrace (who recently had arrived from Harvard where he worked with B. F. (Fred) Skinner). Two years later John A. (Tony) Nevin arrived (with whom I completed my studies, following the unexpected death of Professor Cumming). Since I was interested in visual psychophysics as well as animal learning, much of my graduate work combined these fields in animal psychophysical studies with encouragement and support from these professors at Columbia.

1. Functional relationships for investigating psychophysical processes

At Columbia I took courses in visual psychophysics from Professor Graham who had just published his seminal book: Vision and Visual Psychophysics. Psychophysics was to me the embodiment of systematically varying critical variables to reveal functional relationships and identify lawful relationships. An example of the type of functional relationships I am referring to is shown in Fig. 3. This example serves as a prototype for the approach that, I as a second year graduate student, wanted to bring to the investigation of comparative cognitive processes. Fig. 3a shows the required intensity of a visual stimulus presented for a fixed amount of time to the fovea for seeing. The left-hand (horizontal) limb of the function shows the lawful relationship: log luminance multiplied by time is constant—up to the 'knee' of the function. "In a word" (as Professor Graham was fond of saying) photoreceptors integrate light intensity that falls upon them up to a certain time (the 'knee' in the function) and is known as Bloch's law (or Bunsen–Roscoe law). After the 'knee' there is a second limb to the function; this second limb shows log luminance being constant.

Among aspects that impressed me about those functional relationships shown in Fig. 3 was my realization that if researchers had



Fig. 2. Professor Graham and graduate sponsors Bill Cumming and Tony Nevin at Columbia University. Arrow indicates combining these fields for animal psychophysical studies.



Fig. 3. (a) Functional relationships for an example from human psychophysical detection of light in the fovea, (b) the periphery, (c) and contrast changes from different backgrounds. These functional relationships show that the product of luminance and time was constant for the left-hand 'limb' and luminance was constant for the right-hand 'limb' in each condition.

studied only one or two points (e.g., the two points with an asterisk in Fig. 3a), it is unlikely that these functional relationships would have been discovered. These functional relationships for the fovea extend to peripheral vision (Fig. 3b) and to different levels of adaptation (Fig. 3c), showing that they are basic relationships for seeing and determining vision and how photoreceptors work.

1.1. Pigeon color-vision psychophysics

As an enthusiastic graduate student, I wanted to generate functional relationships for color discrimination (wavelength or hue) of a nonhuman animal, the pigeon (the primary laboratory animal at Columbia), over a substantial range of the pigeon's visible spectrum. And I wanted to compare these discrimination results to the pigeon's category hue boundaries to see if similar relationships would be shown as had been shown for humans. At that time, all psychophysical experiments were conducted with hardware optical systems (e.g., lenses, mirrors etc.). I took an optical engineering graduate course at Columbia, designed my optical apparatus, came up with an inexpensive method to produce controllable narrow-band wavelengths of light (Wright, 1972a), and machined the hardware to mount the lenses, mirrors, light source, and interference filters to control wavelengths (Fig. 4a).

The procedure was a *same/different* task. The pigeons viewed a split field as shown in Fig. 4b. A standard or reference wavelength was presented on the left half of the field. A comparison wavelength was presented on the right half of the split field. On trials where the comparison wavelength was equal to the standard wavelength (16.7% of the trials), the correct response was to peck the left ("same") response key. On trials where the comparison wavelength

was different from the standard wavelength, the correct response was to peck the right ("different") key. Five wavelength differences and a same-wavelength condition were intermixed within each 600-trial session. The probability of reinforcement (mixed grain access) was set to 0.40 for correct responses, but varied systematically across sessions to generate receiver operating characteristics (ROCs).

Two hypothetical ROC functions are shown in Fig. 4c plotted on the unit square (linear % correct) with Hits being correct identifications of wavelength (color) differences and False Alarms being incorrect identifications (different) when there was no wavelength difference (same). Each session produced one point on each of five ROCs, one point for each of the five wavelength differences tested. By plotting the ROCs on *z*-score (standard deviation) scales, they become linear (as they should according to signal detection theory) shown in Fig. 4d.

Fig. 5a and b shows an example from one (out of 4) pigeon, at one (out of 20) reference wavelength (539.8 nm). Only a few nanometers separated the comparison wavelengths from this standard wavelength, demonstrating that hue discrimination was excellent at this spectral point. Linear ROCs of unit slope have the same value of discriminability (d') at all points along the ROC function, but any difference from unit slope will yield different measures of d'. Therefore, d' was computed at the point where the ROC crossed the negative diagonal, a point of neutral bias as shown in Fig. 5b where the proportion of Hits and Correct Rejections are equal (Green and Swets, 1966). These d' values for equal bias (sometimes referred to as d'_s) were computed for the 5 wavelength differences, at each of the 20 reference wavelengths, for the 4 pigeons in this study (Wright, 1972b).



Fig. 4. (a) Top view of apparatus and (b) procedure for studying pigeon color (hue) discrimination including a split field of two wavelengths of light, same/different 'report' pecking keys, and variable reinforcement probability (for correct responses) to generate receiver operating characteristics (ROCs) on (c) linear and (d) z-score axes.



Fig. 5. (a) Computation of *d'* values according to signal detection theory for examples (b) of equal-bias points (circled) from bird 287's linear ROCs on *z*-score axes. (c) Psychometric functions of *d'* as a function of wave number (reciprocal of wavelength) difference for 9 reference wavelengths and examples (reference wavelengths of 649.7 and 620.4 nm) for calculating the wave-number differences for the performance criterion of *d'* = 2.0.



Fig. 6. (a) All 20 psychometric functions for bird 287. (b) Mean psychophysical hue discrimination functions for the group of 4 pigeons at 9 d' criteria values showing best discrimination at spectral points of 600, 540, and 500 nm.

The resulting d' values were then plotted as a function of wavelength difference between the standard and comparison wavelengths as shown for the 9 standard-wavelength examples in Fig. 5c. The psychometric functions shown in Fig. 5c are approximately linear with only small to negligible intercepts and the slopes of these functions change systematically with the standard wavelength. These psychometric functions shown in Fig. 5c can be thought of as second-order functional relationships because they were derived from the ROCs (first-order functional relationships, Fig. 5a). Moreover, from these psychometric functions, psychophysical functions were derived in order to more directly compare the pigeon's hue discrimination. For example, notice the steep slope (i.e., sharply rising accuracy) as the wavelength difference increases for the function on the middle right (600.1 nm), compared to the shallow rise in accuracy for the upper left function (649.7 nm). The psychometric functions were intersected at fixed performance values (criteria) to compare hue discrimination across the pigeon's visible spectrum. Two examples are shown in Fig. 5c for a performance criterion of d' = 2.0. For reference wavelengths of 649.7 nm and 620.4 nm, horizontal lines are drawn at d' = 2.0. At the point where they intersect the psychometric functions specifies the wave-number difference (wave-number is the reciprocal of wavelength) for this performance criterion. By repeating this computation at all 20 of the standard (reference) wavelengths tested resulted in a psychophysical function showing how hue discrimination changes across (that portion of) the visible spectrum. And, of course, by using multiple d' criteria–a matrix of psychophysical functions was produced.

The entire set of 20 psychometric functions for pigeon 287 is shown in Fig. 6a. Those from the other three pigeons were similar (Wright, 1972b, 1978). Mean wave-number differences for the 4 subjects at nine different performance (d') criteria are shown as a matrix of psychophysical functions in Fig. 6b (Wright, 1974). Dips in the psychophysical functions show points of good hue discrimination. These psychophysical functions can be thought of as 3rd order functional relationships because they are derived from the 2nd order psychometric functional relationships (e.g., Fig. 6a). Notice the pronounced dip at 600 nm. At this spectral point, the pigeon's hue discrimination is much better than the human's. (Indeed, I could not see hue differences at this spectral point that the pigeons were regularly discriminating and therefore had to check the dials to make sure the apparatus was working properly before each test session.)

Points of good hue discrimination occur at wavelengths where there is a transition between hues (e.g., blue to green for humans). Fig. 7a (top left) shows transitions between human hues, identified by human color naming. These points of human hue transition line up with the dips in the human wavelength discrimination function (Fig. 7a, bottom left). The overlaid wavelength spectrum, as seen by humans, emphasizes that it is the points of hue transition that produce the best hue discrimination. Using these relationships from human color vision, I wanted to test for similar functional relationships with pigeons. I and my sponsor Bill Cumming had developed a color-naming procedure for pigeons using a matching-to-sample procedure with 3 horizontally aligned pecking keys (Wright and Cumming, 1971). Pigeons learned matching to sample with three different wavelength stimuli and then were transfer tested with intermediate test wavelengths as samples on the center pecking key. The idea was to have the pigeons identify which of two (side key) training wavelengths/colors was more like the test color (center key). At test wavelengths where choices switched from one training wavelength to the other were spectral points of hue transition for the pigeon. These hue-switch points remained invariant over changes in the three training wavelengths revealing two spectral points (542 and 598 nm) of pigeon hue transition (Fig. 7b). These points of hue transition line up with two of the minima of



Fig. 7. (a) Human color naming showing transitions between hues corresponding to points (dips) of best human hue discrimination. (b) Pigeon color naming from a matching-to-sample wavelength generalization experiment showing the correspondence of transitions between three of the pigeon hues to two of the points of best pigeon hue discrimination.

the pigeon's hue discrimination function, a similar relationship as shown for humans. This comparison of color naming and hue discrimination for pigeons can be thought of as 4th order functional relationship (1st order—ROCs; 2nd order—psychometric functions of *d'* vs. wave-number difference; 3rd order—psychophysical functions of wave-number difference for different *d'* criteria; 4th order—hue-discrimination and color-naming comparisons).

Taken together these functional relationships show qualitative similarity between color naming and hue discrimination for both pigeons and humans. But there are quantitative differences; the spectral points of hue transition for pigeons are different than those for humans reflecting differences in photopigments and photoreceptors between these two species.

2. Functional relationships for investigating learning

These functional relationships for pigeon color vision were a prototype for my studies in learning. The logic is similar. But how to generate functional relationships for learning, seemed somewhat less obvious than for psychophysics. Take the issue of animal concept learning. Researchers for years had trained pigeons in matching-to-sample tasks and then tested them for transfer to novel colors. But these studies yielded little or no evidence for abstract-concept learning, prompting comparativecognition theorists to conclude that pigeons did not have the cognitive ability to learn abstract concepts (D'Amato et al., 1985; Premack, 1983; Thomas, 1980, 1996; Thompson, 1995). None of these "failure-to-find" studies, however, had generated functional relationships. From the advantage point of hindsight, this is an issue not unlike those encountered in the early stages of visual psychophysics. One first has to identify some critical variable (or variables) before mapping out a functional relationship (cf., Kamil, 1988)—which brings us to cognitive processes and the topic of functional relationships for concept learning (and memory processing to be discussed later).

2.1. Matching-to-sample learning: pigeons

A matching-to-sample study serves as an example for what a functional relationship may tell us about learning (Wright, 1997). Four groups of pigeons made either 0, 1, 10, or 20 responses to the sample stimulus (cartoon) prior to being presented with two comparison/choice cartoons (see Fig. 8). Other training displays are shown at the bottom of Fig. 8 for one of the two subgroups of each response group. (The other subgroup had right and left positions of comparison cartoons reversed.) A peck to the picture that matched the sample was followed by mixed grain being placed on top of the chosen picture (Fig. 9a). The cartoon pictures were presented on the floor of the test chamber by tipping the video monitor on its back so that the screen pointed up, the rationale being that placing grain reinforcement on top of the correct picture might enhance learning.

All groups learned the task to a high accuracy levels shown in each left-hand (unfilled) histogram for the four groups in Fig. 10. The critical difference between groups was the transfer to novel stimuli shown by the right-hand (dark filled) histograms. (There



Fig. 8. Matching-to-sample procedure for 4 groups (4 pigeons/group) with different sample-response requirements to test item-specific and relational learning with combinations of duck, apple, and grape cartoons. Other training displays for one of the two subgroups are shown at the bottom of the figure.

were five transfer sessions, each contained 84 training trials and 10 transfer trials; transfer-trial examples are shown in Fig. 9b.) The two groups on the left side of Fig. 10 (0-, 1-response groups) showed little or no novel-stimulus transfer. Therefore, if this experiment had been conducted only with these two groups (one required sample response had been typical for most pigeon matching-to-sample experiments), then the conclusion might have been that pigeons do not learn the concept of matching and instead learn the matching task item specifically. (Configural learning, a type of item-specific

learning, was shown by the lack of transfer to training displays of the other subgroup, see Wright, 1997 for further details).

Nevertheless, by including a 10-response group, partial novelstimulus transfer was shown for pigeons (Fig. 10). Had this group been the only group trained and tested, then the conclusion might have been that pigeons can only partially learn the abstract concept of matching. Yet, by adding still another group, the 20-response group, it became clear that pigeons do indeed have the cognitive ability to fully learn the abstract concept of matching. For the



Fig. 9. (a) Apparatus for the matching-to-sample experiment (Fig. 8) with the stimuli projected from the floor and mechanical systems to deliver reinforcement grain on top of correct choice stimuli. (b) Examples of novel-stimulus transfer trials.



Fig. 10. Results from the four different groups of pigeons trained with either 0, 1, 10, or 20 sample pecks prior to the presentation of the choice cartoons. The first bar (unfilled) for each group shows the group baseline training accuracy during transfer testing. The most important transfer result is the right-hand bar for each group which shows novel transfer accuracy and conclusions that can be drawn from the results of each group.

20-response group there was no statistical difference between transfer and baseline performance, showing that they had fully learned the concept. Needless to say, if this had been the only group tested then the conclusion would have been that pigeons learn abstract concepts fully, just like nonhuman primates.

Together, this continuum of sample response effects shows a functional relationship that determines the type of learning—from item-specific learning (i.e., no concept learning) to relational learning and eventually to full concept learning. The partial concept learning of the 10-reponse group is a transition group, likely a blend of these two types of learning (individual subjects showed partial

transfer, not the result of a group average of some with complete transfer and others with no transfer).

2.2. Same/different learning: pigeons, rhesus, capuchins

This learning task was conducted with three species tested with very similar procedures, including the same stimuli and set-size manipulations to produce functional relationships and leading to direct species comparisons (Katz et al., 2002; Katz and Wright, 2006; Wright et al., 2003; Wright and Katz, 2006). The monkeys and pigeons were tested in apparatus appropriate for the species shown



Fig. 11. Rhesus and capuchin monkeys tested in custom aluminum chambers with a juice spout, pellet cup, and template to guide responses. Pigeons tested in custom wooden chamber with a grain hopper and a similar video monitor and touch screen as used with the monkeys.



Fig. 12. Same/different testing procedure with sample touch/peck requirements, "same"/"different" choice responses and examples of the initial 8 training pictures.

in Fig. 11, but they all had the same video monitors, touch screens, and stimuli (equated for visual angle). Pigeons received grain reinforcement for correct responses, whereas monkeys received banana pellets or Tang orange drink. Monkeys, but not pigeons, had a Plexiglas template to guide responses.

Fig. 12 shows the same/different task. Pigeons pecked the sample 20 times (the training condition from the previous experiment that produced concept learning). Half the trials were *same* and half *different*. A "same" response was to the lower picture, and a "different" response was to the white rectangle. Other training pictures



Fig. 13. Training and transfer for monkeys and pigeons with the initial set of 8 pictures of Fig. 12 and the 5 same and 5 different novel transfer trials that were used in the first transfer test session following learning.



Fig. 14. Transfer performance with the training set expanded from 8 to 32 pictures. For pigeons, transfer following training with set sizes expanded from the initial 8-item set but less than 32 items would likely show little or no transfer, whereas monkeys would likely show partial transfer and partial abstract-concept learning.



Fig. 15. Transfer with further expansion of the training set to 64 items and then to 128 item. Pigeons now show partial transfer (and partial concept learning) following training on 128-item set relative to their baseline performance, whereas at this same set size monkeys show transfer equivalent to their baseline performance—and therefore full abstract-concept learning.



Fig. 16. Pigeon transfer with further expansion of the training set to 1024 pictures, resulting in transfer equivalent to their baseline performance and full concept learning—like the monkeys did following training on the 128-item set.

are shown at the bottom of Fig. 12, for the initial training set of 8 pictures. The different species learned the task at similar rates as shown in Fig. 13. notwithstanding an early advantage for many of the pigeons. Following learning (80% criterion), the species were tested for their transfer to novel stimuli. Ten novel stimulus trials were randomly intermixed with 90 training trials (some examples from the first transfer test are shown at the bottom of Fig. 13). None of the species showed any significant transfer or concept learning. Therefore, the training set was expanded to 32 pictures followed by a novel-stimulus transfer test. Following learning with the 32-item training set, some of the species showed some transfer (Fig. 14). If this experiment had been conducted with a training set from the circled region of Fig. 14, the likely conclusion would have been that pigeons do not (possibly cannot) learn a same/different concept, but monkeys do partially learn (transfer less than baseline) this same/different concept.

The training set was then expanded to 64 items followed by learning and testing and then to 128 items. At the 128-item training set, transfer had improved considerably as shown in Fig. 15. If the experiment had been conducted with only this 128-item set size, then the likely conclusion would have been that pigeons can partially learn a same/different concept and that monkeys can fully learn a same/different concept.

Encouraged by these trends, we further expanded the pigeons' training set to 256, 512, and 1024 items as shown in Fig. 16. Following training with these stimulus sets, pigeons also transferred their performance to novel items at a level equivalent to their baseline performance. If the experiment had been conducted only with a 256 (or 512, or 1024) item training set, then the likely conclusion would have been that pigeons, like monkeys, can fully learn a same/different concept and can transfer their performance to novel stimuli equivalent to their baseline performance.

Together these functional relationships show that pigeons, as well as monkeys, have the cognitive ability to learn a same/different abstract concept—a qualitative similarity in learning ability. Nevertheless, it appeared that pigeons required more exemplars

of the rule (i.e., more training pairs) to learn the *same/different* abstract concept than monkeys—a result that would have pointed to a quantitative difference among these species.

Further research, however, showed that altering the pigeon's training resulted in levels of transfer comparable to that shown



Fig. 17. Groups of experimentally naïve pigeons trained initially with either 32 or 64 item sets showing improved transfer which is now equivalent to monkeys trained at these same set sizes with these same items. The results suggest that the previous pigeon groups had detrimental carryover effects from their training with smaller set sizes prior to their transfer at these set sizes.



Fig. 18. (a) Schematic of a 10-item list-memory testing procedure. A monkey hand and arm is shown starting a trial by pressing downward on a lever. List pictures are then sequentially presented on an upper screen. Following a delay, a single test picture is presented on a lower screen. The subject moves the lever to the right, a correct response ("same"), indicating that the test picture was in the list. (Left lever movements would indicate that the test picture was not in the list.) (b) Serial position functions for a monkey (Oscar) and a human tested with the same procedure including stimuli, presentation rates, delays, and response lever. Good memory for the first list items (circled) show recency effects for the monkey and human.

by monkeys. Groups of experimentally naïve pigeons were trained initially on either the 32-item set or the 64-item set (Nakamura et al., 2009). The training stimuli were the same stimuli used to train the pigeons and monkeys at 32- and 64-item set sizes in the prior experiment. The level of transfer following initial training on either the 32- or 64-item sets was equivalent to that for monkeys and was midway between that for capuchin and rhesus monkeys at those same set sizes, as shown in Fig. 17. Despite differences in training history, these results suggest that under some conditions pigeons do transfer at the same level as monkeys and do not need more exemplars of the rule to transfer at these levels.

We suggest that whatever pigeons learn with small training sets, carry over and interfere with learning and transfer with the next larger training set. Transfer can be thought of as a measure of the size of the stimulus domain within which performance is accurate. The stimulus domain for accurate performance seems to grow in some proportion to the training set size. Apparently, the pigeons' domain becomes resistant to growth following learning with small training sets. We have referred to this effect as restricted-domain relational learning (Elmore et al., 2009; Katz et al., 2010; Wright, 2010; Wright and Katz, 2009; Wright and Lickteig, 2010). Restricted-domain carryover effects may be a common property (to a greater or lesser degree) of transfer and generalization—a reluctance to perform outside of one's comfort zone—once something has been learned.

3. Functional relationships for investigating memory processing

3.1. Visual list memory

We began animal list-memory studies after I and my students read (in 1977) Robert Crowder's 1976 seminal book "Principles of

Learning and Memory." We conducted the first nonhuman primate visual list memory study (Sands and Wright, 1980a,b, 1982). In this task shown in Fig. 18a, a rhesus monkey, Oscar, worked in a primate chair facing two vertically aligned back-projection screens. Oscar pushed down on a 3-position level to start a trial. A Carousel projector projected 10 pictures (1-s on, 1-s off) on the upper of two screens. After a 1-s delay a single (test) picture appeared on the lower screen. If the test picture matched one of the list items (as it did in the example in Fig. 18a), then a lever movement to the right was correct and Oscar was reinforced with Tang orange drink. If the test picture did not match any of the list pictures (50% of the time), then a lever movement to the left was correct and was reinforced accordingly. By testing Oscar's memory over many days we were able to determine Oscar's accuracy for different positions in 10-item lists-the serial position function. The serial position function is among the most important cognitive functional relationships and continues to be a test bed for theories of memory (Ebbinghaus, 1902; Glenberg et al., 1983). As shown in Fig. 18b, good memory for the first list items is the primacy effect and good memory for the last list items is the recency effect. This was the first primacy effect for a nonhuman animal. Primacy effects, at the time, were considered to be unique to humans because animals had not shown primacy effects and primacy effects were thought to depend upon rehearsal-a verbally mediated memory strategy instrumental to long-term memory. The recency effect, by contrast, was thought to represent short-term memory. According to the Modal model of memory, items come into short-term memory. If they were sufficiently rehearsed, then they were transferred to long-term memory-resulting in a primacy effect (Atkinson and Shiffrin, 1968). If they were not rehearsed sufficiently, then (according to the model) they were forgotten. Since nonhuman animals were thought incapable of strategic rehearsal control, this monkey primacy effect was unexpected. This



Fig. 19. Examples of two, 4-item list-memory trials with travel slides for testing animal list memory.



Fig. 20. Serial position functions showing primacy and recency effect changes as a function of retention delay for monkeys, pigeons, and humans. (The fourth item is the last list item.) Mean group error bars are shown below each serial position function. Different-trial performance is shown to the right of each serial position function. Animals were tested with "travel pictures" (Fig. 19), and humans were tested with kaleidoscope patterns and kaleidoscope examples are shown on the sides of the figure.



Fig. 21. Top-view schematic of the monkey auditory list-memory procedure. Auditory lists were presented from the front speaker. Following presentation of the list and a delay, a single test sound was played (simultaneously) from both side speakers. A right speaker touch was a "same" response, indicating that the test sound was one of the list sounds. (A left speaker touch would indicate that the test sound was not one of the list sounds.) Other procedure details and names of some of the sounds are shown.

monkey's primacy effect raised issues about the meaning of primacy and recency serial position effects and the role of rehearsal in memory processing.

Further evidence that rehearsal was not necessary for the primacy effect of the serial position function was provided by a series of experiments with the interstimulus interval (ISI) procedure. Increasing ISI had been shown to increase human memory performance (e.g., Intraub, 1980; Proctor, 1983). As the interval between stimuli was progressively increased (e.g., 80-5000 ms) memory performance increased-a remarkable result because more time will have elapsed as the ISI increased which should have resulted in greater decay and forgetting. But just the opposite occurred, suggesting that better memory performance was the result of progressively more rehearsal. We tested list memory of humans and monkeys on the ISI procedure with travel-slide pictures (Cook et al., 1991). Humans (as expected) showed an ISI effect, but monkeys did not, suggesting that the monkeys did not rehearse these pictures in these list-memory tasks. In other human memory tests, we showed that, unlike for travel slides, there was no ISI effect for kaleidoscope pictures-a result suggesting that humans do not spontaneously code or rehearse kaleidoscope pictures. Moreover,



Fig. 22. Mean auditory serial position functions for two monkeys with lists of 6, 8, or 10 sounds.

by teaching names (codes) for 40 kaleidoscope pictures, these same participants then showed ISI effects, and the magnitude of the ISI effect was related to their rehearsal strategy as shown by overt rehearsals and from post-test interviews (Wright et al., 1990a). Also noteworthy was that rehearsal did not affect the magnitude of the primacy effect, but instead improved memory performance for the middle items (the dip in the serial position function). Together these results and findings suggest that many of the same processes that produce serial position effects found in humans can be found in rhesus monkeys—a non-verbal, non-rehearsing animal.

To further explore primacy and recency effects, we used short 4-item memory lists to explore how primacy and recency effects changed with retention delay as well as to accommodate pigeons which had difficulty with list longer than 4 items (e.g., Santiago and Wright, 1984; Wright, 1989, 1998b, 1999a, 2007; Wright et al., 1984, 1985). Lists of four "travel slide" pictures were used to test rhesus monkeys, capuchin monkeys and pigeons (see Fig. 19). Lists of four kaleidoscope pictures were used to test humans to avoid ceiling effects (and also somewhat leveled the "playing" field to the animals).

The serial position functions for the four species are shown in Fig. 20. The form of the serial position function changed systematically with retention interval. At the shortest delay, the serial position function was upward sloping, showing virtually pure recency performance. As the delay was increased, a primacy effect appeared, giving the function its characteristic U-shape. At the longest delays, the recency effect had dropped out, and the serial position function was downward sloping, showing virtually pure primacy performance. These time-course changes for each species can be thought of as 2nd order functional relationships (with the serial position function being the 1st order functional relationship). All four species showed similar changes and a trend toward a primacy effect as time increased. Indeed, Endel Tulving, labeled this trend "The law of Primacy" in a festschrift book honoring Gordon Bower (Gluck et al., 2008). The same pattern of changes for the four species reveals a qualitative similarity and can be thought of as a 3rd order functional relationship (a comparison among 2nd order functional relationships). But there was a time-course



Fig. 23. Comparison of auditory and visual 4-item serial position functions for rhesus monkeys tested at the same retention delays. The auditory and visual serial position functions are opposite in form and are shown to change in opposite ways with changes in delay.

difference among the species. The changes took about 30 s for rhesus and capuchin monkeys, 10 s for pigeons, and 100 s for humans. Different time courses may represent a quantitative difference in memory processing among species.

Together these systematic serial position function changes constrain possible explanations. Consider the consistent result that memory for the first list item improves with retention delay. This is just opposite to the typical notion that memory is supposed to decay with time.

3.2. Auditory list memory

Changes in the visual serial position functions with delay raised the issue as to whether memory in other modalities would show similar changes. For example, would auditory memory reveal similar changes with delay? We therefore embarked upon training rhesus monkeys in an auditory list memory task, despite other researchers having been unsuccessful in this endeavor. We too were unsuccessful-for more than 2 years-but eventually succeeded with a new procedure that required the monkeys to touch the sound source-the speaker. Copper screens were placed in front of the speakers to record touches (via low impedance CMOS circuits). By touching the stimulus, they learned rapidly. We even took advantage of a 'built in' fading procedure, where initially the test sound was played from the correct side speaker only. In just a few trials, the monkeys were touching the speaker from which the sound came. We then gradually increased the sound coming from the incorrect speaker. To our surprise, the monkeys learned the basic task in a matter of several weeks. They also showed full transfer to novel sounds (equivalent to baseline performance) and abstract-concept learning (Wright et al., 1990b)—a feat that was judged to be beyond the cognitive capabilities of rhesus monkeys (e.g., D'Amato et al., 1985; Premack, 1983; Thomas, 1980).

The auditory list memory task is shown in Fig. 21. Monkeys touched the center speaker producing a list of four sounds. Following a delay, a test sound was played simultaneously from *both* side speakers. If the test sound matched one of the list sounds, then the correct touch response was to the right-side speaker (*same* response). If the test sound was different from all the list sounds, then the correct touch response was a touch to the left-side speaker (*different* response). Natural and environmental sounds were used, selected from a 520-item set (Wright and Rivera, 1997).

Auditory serial position functions for 6, 8, and 10 sound lists are shown in Fig. 22. These auditory serial position functions are "U" shaped, not unlike Oscar's visual 10-item function (Fig. 18). One difference is that these auditory serial position functions show an emphasized primacy effect compared to the previously shown visual serial position function. This auditory emphasis on primacy effects will be somewhat more apparent in the next study with shorter auditory lists and delay manipulations similar to those used with visual 4-item lists.

Fig. 23 shows the mean performance of two rhesus monkeys in the auditory 4-item list memory task with delays of 0, 1, 2, 10, 20, and 30 s, along with the rhesus visual 4-item list memory tested at the same delays for comparison. At short delays, the serial position functions were downward sloping, showing virtually pure primacy performance. As the delay increased, a recency effect appeared and grew in strength, giving the function its characteristic U-shape. At longer delays, the primacy effect dropped out, and the serial position function was upward sloping showing virtually pure



Fig. 24. Single-item auditory memory and 4-item auditory list memory with performance for single items and 4th (last) list items circled at 0-, 1-, and 2-s delays to emphasize differences, in spite of the similar events (delay, test) following these items.

recency performance. These opposite shaped serial position functions for auditory memory were a complete surprise to us and were replicated (at the insistence of the Editor) in five additional experiments (Wright, 1998a). Moreover, six further experiments showed the same general form of the results under somewhat different conditions, including with intermixed and blocked delays and after a layoff of 4 years (Wright, 1998b, 1999b, 2002, 2007).

Only by comparing sets of functional relationships for auditory and visual memory were these modality differences apparent. One can think of auditory and visual serial-position-function comparisons as a 3rd order functional relationship (with the 1st order being individual serial position functions, and the 2nd order being changes in the serial position functions with delay).

The importance of different shaped serial position functions for visual and auditory memory is unclear, but may be related to associative-learning evidence showing that visual stimuli are more easily associated with food getting and auditory stimuli are more easily associated with danger avoidance (e.g., Shapiro et al., 1980). How visual memory might be adapted to food getting, is that if an animal has had success foraging for food in distinctive patch, then it should remember to feed in a similar patch (visual recency, short delay). On the other hand, if it feeds in a depleting patch (e.g., berries that ripen in the morning and are depleted as the day wears on), then it will need to remember to go to this patch first thing in the morning after an overnight delay (visual primacy, long delay). For auditory memory, if an animal hears a danger sound, then it will need to remember the starting point of the sound (auditory primacy, short delay) to determine whether the sound (e.g., one made by a predator) is coming toward it and in which direction to escape. On the other hand, if an animal hears a danger sound and the sound stops, then it will need to remember where the sound was last heard (auditory recency, long delay) in order to avoid the spot where a potential predator might be hiding in wait.

To further explore processes that might be responsible for these time dependent auditory serial position function changes, we focused on retrieval inhibition of the subject's memory for the list items at test. In one test, we tested single item memory and compared this performance to memory performance of the fourth list item (Wright and Roediger, 2003). The logic of this experiment was that performance with a single item ought to be just like memory performance for the fourth list item, but with the first three items removed. That is, the memory-item presentation, delay time, and test presentation would be identical for the fourth list item and a single item. Fig. 24 shows that at the three shortest delays there is good memory for single items, but relatively poor memory for the 4th list item. Fig. 25 shows this performance "gap" which is the result of the first three list items inhibiting (proactively) retrieval of memory of the fourth item at these short delays. At the two longest delays, there is also a performance "gap" between single item performance and first item list memory performance (although the comparison is less direct than for the fourth list item). This gap is likely the result of the last three list items retroactively inhibiting retrieval memory of the first item at these long delays. As was true in the prior experiments, only by comparing functional relationships across a substantial portion of the manipulable range was it possible to observe the proactive and retroactive inhibitory effects among auditory list items on rhesus monkey auditory memory.



Fig. 25. Left: Single-item auditory memory compared to 4th (last) item auditory list memory showing the 'gap' in accuracy produced by proactive interference from the previous 3 list items on retrieval memory of the 4th item. Right: Single-item auditory memory compared to first-item of a 4-item auditory list showing the 'gap' in accuracy likely produced by retroactive interference from the last 3 list items on retrieval memory of the 1st item at longer delays.

3.3. Proactive interference in visual memory

In same/different tasks, proactive interference (PI) occurs when previously seen sample pictures are later re-presented as test pictures on trials with nonmatching sample pictures (i.e., *different* trials). Having seen the test picture before, maybe in just in the previous trial, tends to create confusion as to whether this picture was the sample picture in the current trial or in some previous trial. Proactive interference is endemic to all tasks where stimuli are repeated (Keppel and Underwood, 1962). If small sets of stimuli are used, then by necessity, the stimuli will be repeated from trial to trial with Pl growing and eventually saturating. Therefore, investigations of proactive interference need to be conducted with trial-unique stimuli to minimize repetition and interference, thereby allowing effects of specifically placed interfering stimuli to be evaluated.

We conducted a PI test with a rhesus monkey, Oscar, performing the previously described 10-item visual list memory task (Fig. 18). Interfering stimuli were placed 1, 2, 3, 4, 5, or 6 trials prior to the test (Fig. 26). On the PI test, the test item was different from the 10 items in the "current" list (i.e., the correct response was "different"). Having seen the test item before, however, tended to create confusion and increase the chances that Oscar would report "same"-that the test item was in the list on the test trial. When the interfering item was in the immediately preceding trial, there was a large 29% interference effect. Even when the interfering stimulus was 6 trials prior (as many as 60 items before the test), there was still a substantial 10% interference effect (Wright et al., 1986). This PI functional relationship shows how far back (e.g., 60 items) items can interfere with memory performance and consequently how far back they are remembered. If an item is not remembered, then it cannot interfere with later performance.

A single-item memory task, offers a somewhat simpler test of PI. We tested pigeons in a single-item memory task shown in Fig. 27a for interfering stimuli presented 1–16 trials previous to the test (*n*). The task was *same/different*, similar to the one described previously (Fig. 12). Pigeons pecked the sample stimulus 20 times, followed by a delay (1 s or 10 s, in a block design), a test stimulus and white rectangle, choice response, and a 15-s intertrial interval. Baseline-trial stimuli were selected (without replacement) from a 1024 stimulus set (Wright et al., 2012). The results show that even with a 1-s delay, there was considerable PI and that this PI dissipated as a function of increased trial separation. Notice that with a 10-s delay, there is a larger 47% PI effect when the interfering stimulus in the immediately preceding trial. Here too, the PI effect dissipates with increasing trial separation, but interference remains substantially greater than with the 1-s delay.

Greater interference at a longer 10-s delay than at the shorter 1s delay is counterintuitive. With the 10-s delay, interfering stimuli were encountered more distantly in the past than with the 1-s delay (>200 s more distantly for n - 16). More distantly in the past should, according to models of decay or limited capacity, translate to more forgetting and therefore less interference. But just the opposite occurred. We explained this counterintuitive finding using a signal detection theory model showing that interference depended on time ratios: Log time to the current trial sample divided by log



Fig. 26. Proactive interference for a monkey accurately performing a 10-item list memory task (see Fig. 18). On interference test trials, the test picture matched a list picture from a trial seen 1–6 trials previous but differed from all pictures in the current trial. Proactive interference decreases with trial separation showing a proactive interference function.



Fig. 27. (a) Example of two trials from a proactive interference test with pigeons where the interfering stimulus was presented on the preceding trial (n - 1) and repeated as the test stimulus on trial n. (b) Signal detection theory model of elapsed time: Log time to the sample on the current (test) trial $(\log T_c)$ and log time to the interfering sample $(\log T_l)$. (c) Percentage correct performance for 1-s and 10-s delays, and model fits (1 sigma bands) based on the time ratio $\log (T_c/T_l)$ —see text for further explanation.

time to the interfering sample (Fig. 27b). The model was fit simultaneously to both PI functions using the same parameters (bias and maximum accuracy) for each pigeon. The model fit shown in Fig. 27c accounts for 95% of the variance (including the no-PI condition). One implication of this critical time ratio is that the data cannot be explained by "familiarity" models, in which the subject simply reports whether the test stimulus was or was not seen before, including models based on decaying familiarity. According to such models, performance would depend only on the absolute time to



Fig. 28. (a) Examples of two change-detection trials, one with 4 and the other with 6 different clip-art objects, used to test human and monkey short-term memory. (b) Changes in percent correct as a function of display size (number of memory items) for colored objects and clip art objects. (c) Signal detection theory model for detecting (d') which of two test objects was changed. (d) Good model fits to d' plotted as a function of the inverse power law of display size—as predicted from signal detection theory.

b

a

Memory and Mind

A Festschrift for Gordon H. Bower Jos Jony Wright The first student is above the nest menorable and your second hos made the so very proved you. Edited by Gosdon Mark A. Gluck (1/17/07 John R. Anderson Stephen M. Kosslyn

Fig. 29. (a) Gordon Bower receiving the National Medal of Science from President Bush 2007. (b) Note by Gordon Bower on the flyleaf of my copy of his festschrift book "Memory and Mind" (notice the reference to "The Law of Primacy").

the interfering stimulus, not the critical time ratio shown here. Another implication is that time-outs following incorrect responses (popular in the training of animal subjects) should hasten learning by reducing proactive interference, in addition to any effect of delaying the next opportunity for reinforcement—a popular explanation in animal learning.

Of course, none of these findings or conclusions would have been possible without producing functional relationships (i.e., PI functions) for these different delays and showing that a model based upon time ratios could account for the results.

3.4. Functional relationships for short-term memory processes: multiple item displays

Change detection is a popular procedure to study visual short-term memory (VSTM). Change detection is well suited to investigating animal as well as human short-term memory because many memory objects can be presented within the time period of VSTM and change detection appears to be independent of verbal strategies, labeling, or rehearsal.

We compared VSTM of humans and monkeys in similar change detection tasks (Elmore et al., 2011). Two examples trials are shown in Fig. 28a where clip-art objects (from a set of 976) were presented simultaneously in a sample array. Following a retention delay, one of two test objects was changed and the subjects had to touch the changed object. (Monkeys received juice or pellet reward for correct choices and humans received a tone and green light in the testing room.) The number of memory items varied from trial to trial (monkeys: 2-6 items, humans, 2-10 items). Percent correct performance is shown in Fig. 28b and varies as a function of display size for colors (shown in blue and red) and clip art (shown in green and brown). These functional relationships (Fig. 28b) are descriptive accounts which do not specify how memory should vary with display size or how the brain might produce these memory results. By contrast, we proposed a continuous-resource account that stipulates a noisy VSTM (Fig. 28c) and discrimination (d' from signal detection theory) that is predicted to vary as the inverse power law of display size, as confirmed by the good fits $(R^2s > 0.87)$ shown in Fig. 28d (Bays and Husain, 2008; Elmore et al., 2011; Wilken and Ma, 2004). This continuous-resource model accounts for VSTM of both species better than traditional fixedcapacity models (e.g., human capacity: 4 ± 1 ; Cowan, 2001, 2005). Fixed capacity models do not have noisy memory (storage is all or none—a likely invalid assumption given the noisy, probabilistic nature of the nervous system), and predict that capacity will be the same at all display sizes, except for display sizes less than the capacity limit where performance should be perfect (100% correct). Neither requirement was met by humans in this study (nor in some other studies, e.g., Alvarez and Cavanagh, 2004; Eng et al., 2005). Moreover, the computed fixed capacity for monkeys was less than 1 item—unlikely to be a valid conclusion given the species' survival and accurate 10- and 20-item list memory performance by monkey Oscar (see Fig. 18 for 10-item performance and Sands and Wright, 1980a,b). None of these comparisons would have been possible without testing functional relationships for humans and monkeys across a substantial range of their short-term memory for multiple objects.

4. Concluding remarks

This article is organized around a common theme that early on became an integral part of the research-the generation of functional relationships-which my collaborators and I have conducted over the past four decades. More than 17 different studies are presented and discussed showing that functional relationships-generated by varying critical parameters over a substantial portion of their manipulable range-are essential for determining how cognition works. Moreover, the tricky endeavor of comparing cognitive processes across species-comparative cognition-perhaps is most dependent upon functional relationships. Different species need to be tested with the same stimuli and parameters (e.g., viewing time, ISI, delay, visual angle etc.) in order to make direct species comparisons and be effective in claiming cognitive similarities and/or differences among those species. But some conditions often have to vary (e.g., reinforcement type: juice vs. grain, response type: touch vs. peck, testing chamber: large vs. small, etc.). Most important, is whether the task and experimental arrangement are conducive to accurate performance by each of the species. The rub is that seldom will the conditions be equally amenable to accurate performance by the different species, producing uncertainty whether simple accuracy-level differences represent functional cognitive differences or reactions to arbitrary parameter choices. This is where functional relationships come into play. By systematically varying critical parameters (e.g., number of display items to-be-remembered), functional relationships provide a means to effectively compare species despite accuracy level differences (e.g., Fig. 28). Some functional relationships compared in this article suggest qualitative similarities in visual processing, learning, and memory, while at the same time pointing to quantitative differences. Comparing functional relationships (like those shown in this article) in conjunction with studies of the neural basis of this behavior should, in my opinion, be able to provide definitive evidence about cognitive mechanisms and strong implications about their evolution.

Of course, none of what is presented here would have been possible without the many talented collaborators I have been privileged to work with and continuing support from my sponsors. I would be remiss if I did not mention the continuing support from my graduate sponsor, Tony Nevin, graduate committee member, Herb Terrace, and my undergraduate sponsor, Gordon Bower. Several years ago at a celebration for Gordon receiving the 2007 National Medal of Science Award (Fig. 29a), Gordon expressed his support in the fly leaf of my copy of his then-just-published festschrift book (Fig. 29b).

Acknowledgements

Some of the research and preparation of this article was supported by NIH grants R01MH072616 and R01MH091038. I am grateful for the helpful comments and suggestions by Jeff Katz and Jon Crystal (Action Editor).

References

- Alvarez, G.A., Cavanagh, P., 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol. Sci. 15, 106–111.
- Atkinson, R.C., Shiffrin, R.M., 1968. Human memory: a proposed system and its control processes. In: Spence, K.W., Spence, J.T. (Eds.), The Psychology of Learning and Motivation. vol. 2. Academic Press. New York. pp. 89–105.
- Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. Science 321, 851–854.
- Bower, G.H., 1962. An association model for response and training variables in paired-associate learning, Psychol. Rev. 69, 34–53.
- Bower, G.H., 2011. Neal E. Miller. Proc. Am. Philos. Soc. 155, 358-365.
- Cook, R.G., Wright, A.A., Sands, S.F., 1991. Interstimulus interval and viewing time effects in monkey list memory. Anim. Learn. Behav. 19. 153–163.
- Cowan, N., 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–114.
- Cowan, N., 2005. Working Memory Capacity. Psychology Press, New York.
- Crowder, R.G., 1976. Principles of learning and memory. Erlbaum, Hillsdale, NJ.
- D'Amato, M.R., Salmon, D.P., Colombo, M., 1985. Extent and limits of the matching concept in monkeys (*Cebus apella*). J. Exp. Psychol. Anim. Behav. Process. 11, 35–51.
- Ebbinghaus, H.E., 1902. Grundzuge der Psychologie (Basic Psychology). Von Veit, Leipzig.
- Elmore, L.C., Ma, W.J., Magnotti, J.F., Leising, K.K., Passaro, A.D., Katz, J.S., Wright, A.A., 2011. Visual short-term memory compared in rhesus monkeys and humans. Curr. Biol. 21, 975–979.
- Elmore, L.C., Wright, A.A., Rivera, J.J., Katz, J.S., 2009. Individual differences: either relational learning or item-specific learning in a *same/different* task. Learn. Behav. 37, 204–213.
- Eng, H.Y., Chen, D., Jiang, Y., 2005. Visual working memory for simple and complex visual stimuli. Psychon. Bull. Rev. 12, 1127–1133.
- Glenberg, A.M., Bradley, M.M., Kraus, T.A., Renzaglia, G.J., 1983. Studies of the longterm recency effect: support for a contextually guided retrieval hypothesis. J. Exp. Psychol. Learn. Mem. Cogn. 9, 231–255.
- Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.
- Gluck, M.A., Anderson, J.R., Kosslyn, S.M., 2008. Memory and Mind: A Festschrift for Gordon H. Bower. Erlbaum, New York.
- Intraub, H., 1980. Presentation rate and the representation of briefly glimpsed pictures in memory. J. Exp. Psychol. Hum. Learn. Mem. 6, 1–12.
- Kamil, A.C., 1988. A synthetic approach to the study of animal intelligence. In: Leger, D.W. (Ed.), Comparative Perspective in Modern Psychology: Nebraska Symposium on Motivation, vol. 35. University of Nebraska Press, Lincoln, NE, pp. 230–257.
- Katz, J.S., Wright, A.A., Bachevalier, J., 2002. Mechanisms of same/different abstractconcept learning by rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 28 (4), 358–368.

- Katz, J.S., Wright, A.A., 2006. Same/different concept learning by pigeons. J. Exp. Psychol. Anim. Behav. Process. 32, 80–86.
- Katz, J.S., Sturz, B.R., Wright, A.A., 2010. Domain is a moving target for relational learning. Behav. Process. 83, 172–175, PMCID: PMC282002.
- Keppel, G., Underwood, B.J., 1962. Proactive inhibition in short-term retention of single items. J. Verbal Learn. Verbal Behav. 1, 153–161.
- Nakamura, T., Wright, A.A., Katz, J.S., Bodily, K.D., Sturz, B.R., 2009. Abstract-concept learning carryover effects from the initial training set in pigeons (*Columba livia*). J. Comp. Psychol. 123, 79–89, PMID: 19196044.
- Premack, D., 1983. Animal cognition. Annu. Rev. Psychol. 34, 351-362.
- Proctor, R.W., 1983. Recognition memory for pictures as a function of poststimulus interval: an empirical clarification of existing literature. J. Exp. Psychol. Learn. Mem. Cogn. 9, 256–262.
- Sands, S.F., Wright, A.A., 1980a. Primate memory: retention of serial list items by a rhesus monkey. Science 209, 938–940.
- Sands, S.F., Wright, A.A., 1980b. Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. J. Exp. Psychol. Anim. Behav. Process. 6, 386–396.
- Sands, S.F., Wright, A.A., 1982. Monkey and human pictorial memory scanning. Science 216, 1333–1334.
- Santiago, H.C., Wright, A.A., 1984. Pigeon memory: same/different concept learning, serial probe recognition acquisition and probe delay effects in the serial position function. J. Exp. Psychol. Anim. Behav. Process. 10, 498–512.
- Thomas, R.K., 1980. Evolution of intelligence: an approach to its assessment. Brain Behav. Evol. 17, 454–472.
- Shapiro, K.L., Jacobs, W.J., LoLordo, V.M., 1980. Stimulus-reinforcer interactions in Pavlovian conditioning of pigeons: Implications for selective associations. Anim. Learn. Behav. 8, 586–594.
- Thomas, R.K., 1996. Investigating cognitive abilities in animals: unrealized potential. Cogn. Brain Res. 3, 157–166.
- Thompson, R.K.R., 1995. Natural and relational concepts in animals. In: Roitblat, H.L., Meyer, J.-A. (Eds.), Comparative Approaches to Cognitive Science. MIT Press, Cambridge, MA, pp. 175–224.
- Tulving, E., 2008. On the law of primacy. In: Gluck, M.A., Anderson, J.R., Kosslyn, S.M. (Eds.), Memory and Mind. Erlbaum, New York, NY, pp. 31–48.
- Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. J. Vision 4, 1120–1135.
- Wright, A.A., 1972a. Construct a monochromator from a single interference filter. J. Exp. Anal. Behav. 18, 61–63.
- Wright, A.A., 1972b. Psychometric and psychophysical hue discrimination functions for the pigeon. Vision Res. 12, 1447–1464.
- Wright, A.A., 1974. Psychometric and psychophysical theory within a framework of response bias. Psychol. Rev. 81, 322–347.
- Wright, A.A., 1978. Construction of equal-hue discriminability scales for the pigeon. J. Exp. Anal. Behav. 29, 261–266.
- Wright, A.A., 1989. Memory processing by pigeons, monkeys, and people. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation, 23. Academic Press, New York, pp. 25–70.
- Wright, A.A., 1997. Concept learning and learning strategies. Psychol. Sci. 8, 119–123.
- Wright, A.A., 1998a. Auditory list memory in rhesus monkeys. Psychol. Sci. 9, 91–98.
- Wright, A.A., 1998b. Auditory and visual serial position functions obey different laws. Psychon. Bull. Rev. 5, 564–584.
- Wright, A.A., 1999a. Visual list memory in capuchin monkeys (*Cebus apella*). J. Comp. Psychol. 113, 74–80.
- Wright, A.A., 1999b. Auditory list memory and interference in monkeys. J. Exp. Psychol. Anim. Behav. Process. 25, 284–296.
- Wright, A.A., 2002. Monkey auditory list memory: tests with mixed and blocked retention delays. Anim. Learn. Behav. 30, 158–164.
- Wright, A.A., 2007. An experimental analysis of memory processing. J. Exp. Anal. Behav. 88, 405–433.
- Wright, A.A., 2010. Functional relationships for determining similarities and differences in comparative cognition. Behav. Process. 85, 246–251.
- Wright, A.A., Cook, R.G., Rivera, J.J., Shyan, M.R., Neiworth, J.J., Jitsumori, M., 1990a. Naming, rehearsal, and interstimulus interval effects in memory processing. J. Exp. Psychol. Learn. Mem. Cogn. 16, 1043–1059.
- Wright, A.A., Cumming, W.W., 1971. Color naming functions for the pigeon. J. Exp. Anal. Behav. 15, 7–17.
- Wright, A.A., Katz, J.S., 2006. Mechanisms of same/different concept learning in primates and avians. Behav. Process. 72, 234–254.
- Wright, A.A., Katz, J.S., 2009. A case for restricted-domain relational learning. Psychon. Bull. Rev. 16, 907–913, PMID: 19815797.
- Wright, A.A., Katz, J.S., Ma, W., 2012. How to be proactive about interference: lessons from animal memory. Psychol. Sci. 23, 453–458, http://dx.doi.org/10.1177/ 0956797611430096.
- Wright, A.A., Lickteig, M.T., 2010. What is learned when concept learning fails: a theory of restricted-domain relational learning. Learn. Motiv. 4 (4), 273–286.
- Wright, A.A., Rivera, J.J., 1997. Memory of auditory lists by rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 23, 441–449.
- Wright, A.A., Rivera, J.J., Katz, J.S., Bachevalier, J., 2003. Abstract-concept learning and list-memory processing by capuchin and rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 29 (3), 184–198.
- Wright, A.A., Roediger III, H.L., 2003. Interference processes in monkey auditory list memory. Psychon. Bull. Rev. 10 (3), 696–702.

- Wright, A.A., Santiago, H.C., Sands, S.F., 1984. Monkey memory: same/different concept learning, serial probe acquisition, and probe delay effects. J. Exp. Psychol. Anim. Behav. Process. 10, 513–529.
- Wright, A.A., Santiago, H.C., Sands, S.F., Kendrick, D.F., Cook, R.G., 1985. Memory processing of serial lists by pigeons, monkeys, and people. Science 229, 287–289.
- Wright, A.A., Shyan, M.R., Jitsumori, M., 1990b. Auditory same/different concept learning by monkeys. Anim. Learn. Behav. 18, 287–294.
- Wright, A.A., Urcuioli, P.J., Sands, S.F., 1986. Proactive interference in animal memory research. In: Kendrick, D.F., Rilling, M., Denny, R. (Eds.), Theories of Animal Memory. Erlbaum, Hillsdale, NJ, pp. 101–125.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Change detection for the study of object and location memory

L. Caitlin Elmore*, Antony D. Passaro, Anthony A. Wright

The University of Texas Health Science Center at Houston, United States

ARTICLE INFO

Article history: Received 18 July 2012 Received in revised form 31 October 2012 Accepted 5 November 2012

Keywords: Change detection Visual short-term memory Object memory Location memory

ABSTRACT

Seven adult human participants were tested in change detection tasks for object and location memory with large and small sets of four different stimulus types. Blocked tests demonstrated that participants performed similarly in separate object and location tests with matched parameters and displays. In mixed tests, participants were informed that they would be tested with *either* object changes or location changes; surprisingly, they were nearly as accurate remembering both objects and locations as when either was tested alone. By contrast, in the large-set condition, performance was lower than baseline on surprise probe test trials in which participants were tested (on 13% of trials) with the change type opposite to the present block (e.g., location probe trials during the object change block). These probe-test results were further supported by the reduction in probe-baseline differences when tested with small sets (6) of these item types. Small sets required remembering locations and objects to resolve object-location confounds. Together these results show that humans can remember both objects and locations without instruction. © 2012 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, visual short-term memory has been studied in both humans and animals using the change detection task (e.g., Wilken and Ma, 2004; Alvarez and Cavanagh, 2004; Eng et al., 2005; Wright et al., 2010; Elmore et al., 2011; Heyselaar et al., 2011; Elmore et al., 2012). In this task, participants are presented with a display of visual objects, and after a brief retention delay are asked to report either the presence or absence of a change, or the specific item that has changed in a test display. The task has primarily been used to study object memory, and researchers frequently investigate the amount of information (number of objects) remembered or precision of memory (ability to detect signal from noisy representations). However, the change detection task also lends itself to the study of spatial memory (memory for locations). The task allows multiple stimuli to be presented simultaneously in multiple locations for participants to remember. Instead of asking participants to identify a changed object, one can ask participants to identify changes in an object's location following a retention delay.

Using change detection to study memory for locations is also advantageous because location memory can be directly compared to object memory using exactly the same stimuli and task parameters. In addition, it is important to study memory for objects and

E-mail address: lelmore@cns.bcm.edu (L.C. Elmore).

locations concurrently because all objects necessarily occupy a location and locations are marked by the presence (or absence) of objects. In fact, research has indicated that objects and their locations are "bound" together in short-term memory under some conditions (Wheeler and Treisman, 2002). Participants can be asked to store both types of information on every trial. Also, probe tests can be conducted in which participants are instructed to attend to one type of information (e.g., object) and are then probed with unanticipated location change trials to see the extent to which they are storing location information as well, and vice versa.

Frequently, change detection studies are conducted using small sets of stimuli. Consequently a given stimulus will often repeat both within a single trial and across a series of trials. If a stimulus is presented more than once in a given trial, participants will be forced to also attend to the object's location in order to differentiate between identical objects and accurately perform the task. Repetition of stimuli across trials can lead to the buildup of proactive interference which is often detrimental to performance (e.g., Wright et al., 2012; Makovski and Jiang, 2008; Roberts and Grant, 1976) Thus it is also important to investigate the role of set size in short-term memory performance.

The goal of the present study was to directly compare object and location short-term memory using a change detection task. In addition, the study sought to further elucidate the cognitive processing used in a mixed condition, where objects and their locations had to be maintained in memory concurrently. The purpose of this condition was to assess whether memory performance would suffer when the memory load was effectively doubled by requiring participants to memorize both object and location information within the

^{*} Corresponding author at: Department of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, United States. Tel.: +1 713 858 6753.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.002

same trial. Next, in a probe condition, we assessed the role of conscious awareness in the tendency to bind objects and their locations in memory. The probe condition examined whether participants would remember both object and location information when they were instructed that it was only necessary to remember one type of information. Finally for all conditions (object, location, mixed, and probe) we assessed the role of large and small sets of stimuli in change detection memory performance.

2. Methods

2.1. Participants

Seven adult human participants were recruited to participate in this study. They ranged in age from 23 to 28 (mean age 25.6), and there were five females and two males. The participants visited the lab for a total of eight 1-h sessions. The participants were compensated \$10 per 1-h session. All procedures were approved by the University of Texas Health Science Center at Houston Committee for the Protection of Human Subjects.

2.2. Apparatus

The participants were tested in a room with a PC computer. The computer's monitor (17" EIZO) was equipped with an infrared touch-screen (17-in. Unitouch; ELO, Round Rock, TX). The participants were provided feedback by two 25 W light bulbs that were mounted on the wall behind the participants. The green light was illuminated for 1 s following correct responses and the red light was illuminated for 1 s following incorrect responses. The lights were operated by a computer-controlled relay interface (Model PI0-12; Metrabyte, Taunton, MA). Microsoft Visual Basic 6.0 was used to create custom software which created, controlled, and recorded experimental sessions. The monitor was controlled by a video card (ATI graphics adaptor).

2.3. Stimuli

The stimuli were 976 color clip art images, 976 color kaleidoscope images, 256 black and white Kanji characters, and 256 black and white Snodgrass line drawings (Snodgrass and Vanderwart, 1980). A subset of the stimuli is depicted in Fig. 1. The stimuli were randomly presented in 20 possible locations (defined by points on two invisible concentric circles). The stimuli subtended a visual angle of 1.3°.

2.4. Test procedures

The participants completed a total of eight 1-h test sessions of the change detection task. In this task, participants first viewed a sample display of six, eight, or ten stimuli (all from the same category in an individual trial) for 1 s. A black display was then presented for a 1-s delay period. Following the delay participants were presented with the test display which contained two stimuli, one of which matched (object identity and location) a stimulus from the sample display, and one of which had changed (either in object identity or location). The participants' task was to touch the stimulus that had changed. The trial sequence is depicted in Fig. 2. Individual trials were restricted to one stimulus category (e.g., Kanji characters), but stimulus categories were intermixed within the session. Each 1 h test session consisted of two blocks of 180 trials, in which stimulus categories and display sizes (6, 8, 10) were randomly intermixed.

Table 1

Percent	correct	and	S.E.M.s	for	object	and	location	changes	with	large	and	small
stimulu	s sets.											

Stimulus type	Blocked		Mixed					
	Object	Location	Object	Location				
Large stimulus set								
Clip art	76.49 ± 1.61	76.82 ± 0.32	75.93 ± 5.31	74.35 ± 4.79				
K-scopes	66.12 ± 4.75	71.99 ± 2.90	55.02 ± 3.06	68.75 ± 5.69				
Kanji	63.55 ± 1.19	71.67 ± 3.75	65.85 ± 2.28	72.73 ± 0.89				
Snodgrass	77.81 ± 3.58	72.50 ± 2.54	70.39 ± 3.72	73.35 ± 5.59				
Small stimulus set								
Clip art	73.31 ± 4.48	73.09 ± 1.38	79.76 ± 4.62	72.78 ± 3.10				
K-scopes	65.86 ± 5.44	72.98 ± 3.60	54.63 ± 1.88	68.87 ± 3.10				
Kanji	64.69 ± 2.25	69.55 ± 1.39	70.15 ± 6.12	76.34 ± 1.86				
Snodgrass	74.09 ± 1.20	71.10 ± 3.94	72.66 ± 2.78	72.49 ± 2.48				

2.5. Test conditions

The eight test sessions were divided into two groups of four. One group of four sessions belonged to the large set condition, in which all stimuli in all trials were drawn from a large group of 976 (clip art and kaleidoscopes) or 256 (Kanji and Snodgrass) stimuli. The second group of four sessions belonged to the small set condition. In the small set condition, trials were drawn from sets of six stimuli from each category with no more than two repeats for each stimulus in a given sample display. For each condition (small and large set), the four test sessions were divided into eight blocks of 180 trials (two each of object change, location change, probe test, and mixed condition). The order of the blocks tested was counterbalanced. In the object change condition, participants were instructed to memorize the objects in the sample display and look for a change in the object's identity in the test display. In the location change condition, participants were instructed to memorize the objects' locations in the sample display and look for a change in location in the test display.

In the probe condition, for the object-change block of 180 trials, participants were instructed that the trials were in the object change condition and that they should look for changes in object identity. However, 24 probe trials were intermixed in which there was no object change, but rather a change in location of one object. Likewise, in the location block of 180 trials there were 24 probe object change trials intermixed. In both cases, participants were not informed of the probe trials. Lastly, in the mixed condition, object and location trials were randomly intermixed (90 of each per block), and participants were instructed that there could be a change in object identity or location, and that they should therefore try to memorize the objects' identities and locations during the sample display and look for either type of change during the test display.

3. Results

Mean accuracies and standard errors for each stimulus type in the large and small set condition are listed in Table 1.

Fig. 3A displays mean accuracy in each of four trial types (blocked object and location and mixed object and location) for the large set condition. Performance was significantly greater than chance in all trial types (binomial tests, $p_s \le 0.0001$). Fig. 3B displays mean accuracy in the blocked object and location and mixed object and location trial types for the small set condition. Performance was once again significantly greater than chance in all conditions (binomial tests, $p_s \le 0.0001$). A five-way factorial analysis of variance (display size × set size × trial type × stimulus type × condition) was conducted. The ANOVA showed highly significant effects of display size (F(2,975)=50.68, $p=1.15 \times 10^{-21}$), stimulus type (F(3,975)=37.79, $p=4.13 \times 10^{-23}$), trial type (object or location:



Fig. 1. Example stimuli. First row: clip art. Second row: kaleidoscope images. Third row: Kanji characters. Fourth row: Snodgrass line drawings.

F(1,975) = 15.91, $p = 7.15 \times 10^{-5}$), and condition (blocked or mixed: F(1,975) = 6.48, p = 0.01). Participants had higher accuracy with smaller display sizes, and performed better with clip art and Snodgrass stimuli than with Kanji and Kaleidoscopes. See Table 1 for detailed accuracy results. In addition, participants performed more accurately in the location change trials (72.33% correct) than in the object change trials (69.77% correct). Participants were also slightly more accurate overall in the blocked condition (71.61% correct) than in the mixed condition (69.83% correct). There were also significant interactions of display size and stimulus type (F(6,975) = 4.46, p = 0.0002) and of trial type and stimulus type (F(3,975) = 17.25, $p = 6.39 \times 10^{-11}$).

3.1. Probe condition

Fig. 4A and B displays mean performance in the probe condition. In Fig. 4A the large set probe condition is shown and in Fig. 4B the small set probe condition is shown. Participants performed well on baseline trials, both in the object and location conditions. Baseline refers to change type trials that the participants were instructed to perform (e.g., object or location change). Probe trials refer to the change types of which the participants were not instructed (probe trials of the other change type intermixed – e.g., location change trials in an object change block). A four-way factorial ANOVA of display size × set size × trial type × probe condition (baseline or probe) showed a significant effect of display size (F(2,153)=3.77, p=0.03) and probe condition (F(1,153)=21.86, $p=6.4 \times 10^{-6}$). Thus in both the large and small set conditions, participants performed better with smaller display sizes, and accuracy was greater in the baseline trials than in the unanticipated probe trials.

4. Discussion

In the blocked condition, participants performed accurately in both the object and location change trials, although location change performance was significantly higher (72.33% correct vs. 69.77%



Fig. 2. Trial progression in the change detection task. The schematic presented here is representative of the mixed condition in which object and location trials are intermixed.



Fig. 3. (A) Change detection accuracy by trial type in the large set condition. (B) Change detection accuracy by trial type in the small set condition. Error bars represent standard error of the mean.

correct for object change trials). There was also a small but significant difference in overall accuracy between the blocked (71.61% correct) and mixed conditions (69.83% correct). However, this 1.78% difference is much smaller than would be predicted based on the fact that the mixed condition required subjects to store twice as much information on every trial relative to the blocked condition. During the mixed condition, subjects needed to maintain both object and location information for each stimulus in the sample display, while in the blocked condition subjects were only required to maintain one piece of information (either object identity or location) about each stimulus in the sample display. The very small difference in performance between the blocked and mixed conditions suggests that object and location information is processed in parallel, with very little cost to overall performance. Such a small decrement in performance could be attributed to a constraint on attentional resources, rather than a limitation in memory storage, although the experiments conducted here do not allow that question to be directly addressed.

Wheeler and Treisman (2002) showed that focused attention is necessary for the maintenance of binding over time. While one might expect that the association of an object and its context (location) is automatic, both our results and the results of Wheeler and Treisman (2002) suggest that the parallel processing of object and location information may be under conscious control. Probe test trial performance was significantly worse than baseline object and location performance. Thus, when participants were unaware that object and location information should be stored concurrently, they failed to encode both pieces of information and instead stored what they had been instructed to remember (e.g., object or location alone).



Fig. 4. (A) Change detection accuracy by trial type in the large set probe condition. Participants completed two 180-trial blocks, one each of object and location change detection in the large set probe condition. In each block, 24 probe trials of the opposite change type were randomly intermixed. (B) Change detection accuracy by trial type in the small set probe condition. Error bars represent standard error of the mean.

It is also possible that subjects stored object and location information in parallel but were biased by the experimenter's instructions to perform object or location change detection alone. Nevertheless, in interviews following the experimental session, none of the participants reported noticing any "unusual trials" during the probe condition.

The smaller baseline-probe difference in the small-set condition (8.58% difference between baseline and probe vs. 14.9% difference in the large set condition) may have been due to participants having to resolve the ambiguity between object and location when stimuli were repeated in the same display. In the small set condition, it was necessary for participants to attend to both object and location information in order to accurately perform the task. For example, in the object condition, using clip art stimuli as an example, the football helmet might appear as part of the sample display in location 1. After the delay, the dinosaur in location 2 could change to the football helmet. If the subject attended to object information alone, they would not accurately detect the change because the football helmet appears in both the sample and test display (see Fig. 5). Although participants in the small set condition might have been biased toward the block type (e.g., object changes), the need to resolve the object-location confound described above likely facilitated good performance on probe trials, as participants may have actively attended to both types of information. Thus, the results from the small set condition support the idea that the binding of object and location information is under conscious control.



Fig. 5. Example object change trial with object-location confound. The dinosaur in the sample display changes to the football helmet in the test display. The football helmet was present in another position in the sample display, and the subject could be confused by this, thinking that the football helmet was not the changed item because it was present in the sample display.

A similar series of probe tests was conducted with both rhesus monkeys and pigeons trained to perform object change detection (see Elmore et al., 2012). These animals had no experience with location change trials and were tested to see if they would spontaneously transfer their change detection performance to location changes. Interestingly, in this case there was a species difference, as shown in Fig. 6. Monkeys performed as well with the probe location change trials as they did with their baseline object change trials. Similar to the small set probe condition with human subjects, the monkeys were trained and tested with a small set of eight colored circle stimuli such that an optimal strategy would be to attend to both object and location information simultaneously, thereby facilitating their good performance with probe location change trials. Pigeons, however, performed at chance with location changes after being trained to perform object change. Although pigeons were



Fig. 6. Performance by rhesus monkeys and pigeons trained in the object condition and tested with probe location change trials. Error bars represent standard error of the mean. Participants were tested over the course of seven 96-trial session with twelve probe trials randomly intermixed in each session.

also trained with a small set of stimuli they have been shown to be quite poor at transferring their performance to novel types of change (Elmore et al., 2012) so it is not surprising that they did not perform well with the novel probe location change trials. In addition, this is consistent with a species difference in bias. Pigeons are potentially more biased by their training with object changes than the rhesus monkeys, who appear to have stored object and location information concurrently, and readily identified both types of change.

In recent years, much emphasis has been placed on the notion of a slot-like storage system for visual short-term memory (e.g., Luck and Vogel, 1997; Cowan, 2000; Alvarez and Cavanagh, 2004; Eng et al., 2005; Buschman et al., 2011). This work supports the idea of a fixed capacity for visual information (e.g., magic number 4 ± 1). However, if visual memory is limited to specific number of stimuli, how can one explain the findings from comparing our blocked and mixed conditions? If a subject is only able to accurately store four stimuli in the blocked condition due to their limited-capacity slotlike storage system, how is it possible that the subject is able to perform nearly as well when they have twice as much information to store in the mixed condition? One could theorize that object and location information are bound together in memory and that the slots are filled by bound units (one object + its location).

Finite limitations in visual short-term memory have come under scrutiny in other recent work. These studies have supported the continuous-resource model using data from both humans (Wilken and Ma, 2004) and rhesus monkeys (Elmore et al., 2011). Rather than modeling memory as a discrete entity of a few slots, the continuous-resource model states that memory is a continuous resource that can be allocated to many stimuli. Instead of capacity, the model uses d' from signal detection theory as a measure of memory sensitivity. A reduction in resource per stimulus with increasing memory load results in increasing noise in those memory representations. If objects and locations are bound together in memory, perhaps the memory resource is allocated to each bound unit (one object + its location), or allocated globally across the scene as a whole instead of on a stimulus by stimulus basis.

Future experimental and theoretical work should seek to better understand the interplay of object and location memory. Is binding under conscious control? Or are participants simply biased by their expectations of the task? How do models of visual short-term memory account for our findings that participants actively maintain both object and location memory with similar accuracy, but only when instructed (mixed condition) or when it was advantageous (small set probe condition)? In our daily lives, our memory requirements are rarely restricted to the simple case of a small display of visual objects. We need to maintain vast stores of memories of scenes of information in order to navigate our environment. As a result, a cohesive understanding of the intersection of object and location memory is important. In addition, understanding the interplay between object and location memory also has implications for the "what" and "where" of episodic memory (what, when, and where) (e.g., Tulving, 2001, 2002; Clayton and Dickinson, 1998; Clayton et al., 2001, 2003).

Acknowledgements

Research and preparation of this article was supported by NIH grants R01MH072616 and R01MH091038 to A.A.W.

References

- Alvarez, G.A., Cavanagh, P., 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol. Sci. 15, 106–111.
- Buschman, T.J., Siegel, M., Roy, J.E., Miller, E.K., 2011. Neural substrates of cognitive capacity limitations. PNAS 108, 11252–11255.
- Clayton, N.S., Dickinson, A., 1998. Episodic-like memory during cache recovery by scrub jays. Nature 395, 272–274.
- Clayton, N.S., Yu, K.S., Dickinson, A., 2001. Scrub jays (Aphelocoma coerulescens) form integrated memories of the multiple features of caching episodes. J. Exp. Psychol. Anim. Behav. Process. 27, 17–29.

- Clayton, N.S., Bussey, T.J., Dickinson, A., 2003. Can animals recall the past and plan for the future? Nat. Rev. Neurosci. 4, 685–691.
- Cowan, N., 2000. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–185.
- Elmore, L.C., Ma, W.J., Magnotti, J.F., Leising, K.J., Passaro, A.D., Katz, J.S., Wright, A.A., 2011. Visual short-term memory compared in rhesus monkeys and humans. Curr. Biol. 21, 975–979.
- Elmore, L.C., Magnotti, J.F., Katz, J.S., Wright, A.A., 2012. Change detection by rhesus monkeys (*Macaca mulatta*) and pigeons (*Columba livia*). J. Comp. Psychol. 3, 203–212.
- Eng, H.Y., Chen, D., Jiang, Y., 2005. Visual working memory for simple and complex visual stimuli. Psychon. Bull. Rev. 12, 1127–1133.
- Heyselaar, E., Johnston, K., Pare, M., 2011. A change detection approach to study visual working memory of the macaque monkey. J. Vision 11, 1–10.
- Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. Nature 390, 279–281.
- Makovski, T., Jiang, Y.V., 2008. Proactive interference from items previously stored in visual working memory. Mem. Cognit. 36, 43–52.
- Roberts, W.A., Grant, D.S., 1976. Studies of short-term memory in the pigeon using the delayed matching-to-sample procedure. In: Medin, D.L., Roberts, W.A., Davis, R.T. (Eds.), Processes of Animal Memory. Erlbaum, Hillsdale, NJ, pp. 79–112.
- Snodgrass, J.G., Vanderwart, M., 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J. Exp. Psychol. [Hum. Learn]. 6, 174–215.
- Tulving, E., 2001. Episodic memory and common sense: how far apart? Philos. Trans. R. Soc. Lond. B Biol. Sci. 1413, 1505–1515.
- Tulving, E., 2002. Episodic memory: from mind to brain. Annu. Rev. Psychol. 53, 1–25.
- Wheeler, M.E., Treisman, A.M., 2002. Binding in short-term visual memory. J. Exp. Psychol. Gen. 131, 48–64.
- Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. J. Vision 4, 1120–1135.
- Wright, A.A., Katz, J.S., Ma, W.J., 2012. How to be proactive about interference: lessons from animal memory. Psychol. Sci. 23, 453–458.
- Wright, A.A., Katz, J.S., Magnotti, J., Elmore, L.C., Babb, S., Alwin, S., 2010. Testing pigeon memory in a change detection task. Psychon. Bull. Rev. 17, 243–249.

Contents lists available at SciVerse ScienceDirect

Behavioural Processes





John F. Magnotti^{a,b,*}, Adam M. Goodman^a, Thomas A. Daniel^a, L. Caitlin Elmore^{b,c}, Anthony A. Wright^b, Jeffrey S. Katz^a

^a Department of Psychology, Auburn University, Auburn, AL, USA

^b Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX, USA

^c Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Article history: Received 23 July 2012 Received in revised form 8 October 2012 Accepted 15 October 2012

Keywords: Visual short-term memory Visual target search Change detection

ABSTRACT

The study of visual memory has repeatedly shown qualitatively similar visual short-term memory (VSTM) systems between human and many nonhuman species. In studies of human VSTM using change detection, increasing visual object complexity has an inverse effect on accuracy. In the current study, we assessed the functional relationship between visual object complexity and memory performance in visual change detection in pigeons and humans. Visual object complexity was quantified for each object type within each species using visual target search. Change detection performance was inversely related to object complexity in both species, suggesting that pigeon VSTM, like human VSTM, is limited by visual object complexity. Human participants were able to use a verbal-labeling strategy to mitigate some of the effect of visual object complexity, suggesting a qualitative difference in how the two species may solve certain visual discriminations. Considering the visual complexity of novel objects may also help explain previous failures to transfer relational rules to novel visual objects.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Of the many lasting impacts Tony Wright has had on the fields of human and nonhuman memory, his emphasis on estimating functional relationships via broad parametric manipulation is perhaps the most important. His work demonstrating cross-species similarities in list memory (for a review, Wright, 2007) is a striking example of how testing a broad range of the parameter space can reveal surprising functional relationships. We have adopted a similar parametric approach in our recent work investigating visual short-term memory (VSTM) in pigeons, humans, and nonhuman primates (e.g., Elmore et al., 2011, 2012; Wright et al., 2010). The goal of the current study was to look for a functional relationship between object complexity and VSTM in pigeons and compare it to an analogous function from humans.

Humans and pigeons alike rely on VSTM to accomplish a variety of tasks ranging from navigating a crowded environment to locating food; the large degree of qualitative similarity in visual processing is therefore not surprising. Work in our laboratories and others has shown that, like human VSTM, pigeon and nonhuman primate VSTM is sharply limited in terms of available resources (e.g., Elmore et al., 2011; Gibson et al., 2011; Heyselaar et al., 2011) and susceptible to proactive interference (e.g., Wright et al., 2012).

An increasingly common way to study VSTM across species is the change detection procedure. In the version we employ (e.g., Wright et al., 2010), trials begin with the presentation of a sample array containing items from a particular object type (e.g., colored circles, clip art), followed by a brief probe delay. After the probe delay, a probe array is presented containing two items, one of which has changed to another item from the object type. The participant's goal is to choose the item that changed (or equivalently, is "different") from the sample array. Performance in the task requires participants to (a) successfully encode some portion of the sample array, (b) retain this information throughout the probe delay, and (c) determine which item in the probe array was not contained in the sample array.

Although VSTM is often conceptualized as being comprised of a fixed number of objects (e.g., Luck and Vogel, 1997; Zhang and Luck, 2008; but see Bays and Husain, 2008; van den Berg et al., 2012), a consistent finding in studies of human VSTM is a large effect of object complexity (Alvarez and Cavanagh, 2004; Eng et al., 2005). As the objects become more complex, accuracy (and subsequent capacity estimates) significantly decreases. Importantly, the complexity of a given object type can be independently assessed by using a target search task (also known as a visual search task).

Alvarez and Cavanagh (2004) presented participants with alternating blocks of change detection and target search using objects of varying complexity, ranging from colored squares (low complexity)



^{ightarrow} Jon Crystal was the acting editor for this manuscript.

^{*} Corresponding author at: Department of Neurobiology & Anatomy, University of Texas Medical School at Houston, 6431 Fannin Street, TX 77030, United States. *E-mail address:* john.magnotti@auburn.edu (J.F. Magnotti).

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.10.006

to shaded cubes (high complexity). For each object type, search rate slopes were obtained from the target search response time data and paired with the corresponding change detection data. Across 6 object types, search rate explained over 99% of the variance in change detection performance. The authors claimed that this relationship demonstrated that the amount of information stored in VSTM was related not only to the number of objects in memory, but also to the complexity of the objects being stored.

Along similar lines, Eng et al. (2005) related target search and change detection across object types in humans. To rule out encoding-related explanations for the complexity–capacity relationship, participants were given variable amounts of time to study the sample array (including one self-paced condition). Their results showed that the complexity–capacity relationship is weakened (but does not disappear) when the sample array viewing time is increased, suggesting that encoding limitations do play a role in change detection for complex objects. The authors noted, however, that as sample–array viewing time is increased, alternative strategies become available (e.g., chunking, verbal coding) that may change how the objects are processed.

An alternative account offered by Awh et al. (2007) is that the difficulty of *comparing* the sample and probe arrays (rather than just the *encoding* of the sample array) is responsible for the decreased accuracy in change detection with complex objects. They showed participants sample arrays of mixed object types and allowed objects to have cross-category changes (e.g., a Chinese character changing to a cube). The accuracy difference between object types was only significant for within-category changes, despite identical encoding and storage demands. They concluded that the primary driver of the negative relationship between object complexity and accuracy must therefore be the sample/probe array comparison (Awh et al., 2007). Regardless of the exact reason for the decrease in accuracy, if pigeon VSTM is qualitatively similar to human VSTM, then pigeons should show a similar effect of complexity on change detection performance.

We recently assessed the degree to which pigeons transferred performance from color change detection to other domains (Elmore et al., 2012). Although pigeons have previously transferred change detection performance to novel colors in a previous study (Wright et al., 2010), transfer performance on novel shape changes was poor. Even after extensive training with shape training, performance never reached the same level as color change detection performance. Although the primary conclusion concerning restricted-domain relational learning explains the transfer test results (Elmore et al., 2012), the inability of pigeons to learn shape change detection may be related additionally to the increased visual complexity of shapes as compared to colors. Previous work has shown that visual search slopes can provide an independent, quantitative assessment of object complexity that describes decreases in VSTM performance (Alvarez and Cavanagh, 2004; Eng et al., 2005).

1.1. Experimental overview

To test the visual complexity hypothesis, in Experiment 1 we trained 2 birds in a visual target search task using the objects from our previous change detection experiments. Although previous human studies comparing target search and change detection relied on within-subject comparisons, using different pigeons was essential to eliminate carryover effects caused by the opposing natures of the two tasks – our target search task is equivalent to a change detection task with 1 sample item and rewarding responses to the object that matches the sample array (rather than the one that changed). To help ensure our forced-choice task, small-n design, and between-subject manipulation were not uniquely driving any observed relationship between change

detection and target search, Experiment 2 was conducted to obtain target search slopes from humans in the task from Experiment 1. These slopes were then compared with results obtained in the human variant of our forced-choice task (Elmore et al., 2013).

2. Experiment 1

The main purpose of Experiment 1 was to establish target search functions with color, shape, and polygon object types. Search functions provide a quantitative estimate of object complexity that can then be compared with change detection performance from identical object types (cf. Alvarez and Cavanagh, 2004; Eng et al., 2005). The method we developed for target search was therefore designed toward achieving a close match with our previous change detection work.

2.1. Method

2.1.1. Subjects

Two White Carneaux Pigeons (*Columba livia*) with previous experience in a *same/different* procedure served as subjects. Testing was conducted 5 days per week. The pigeons were maintained at 80–85% range of their free-feeding weights, with free access to grit and water in their individual home cages. A14:10-h light:dark cycle was maintained in the room containing individual home cages.

2.1.2. Apparatus

The pigeons were tested in a $(35.9 \text{ cm wide} \times 45.7 \text{ cm} \text{ deep} \times 51.4 \text{ cm high})$ custom wooden chamber. A hopper delivered mixed grain through a cutout centered below a 17-in. Eizo T550 color monitor (resolution of 800×600 pixels). A pressure-fitted infrared touchscreen (Carroll Touch, Round Rock, TX) was used to measure responses. A house light (Chicago Miniature 1829, 24V) was located in the center of the ceiling and was illuminated only during intertrial intervals (ITIs). A rear fan provided ventilation and white noise.

2.1.3. Objects

Object types consisted of 8 colored circles, 8 white shapes, and 8 white irregular polygons, as shown in Fig. 1. One pigeon was initially trained with 72 irregular polygons, but training was switched to the 8 items following poor acquisition, and we show only the 8 used for extensive training. Additionally, because of poor shape acquisition, 4 shapes (the first four from the left in Fig. 1) were used by the end of shape training. Objects were chosen to be identical to those used in previous change detection studies in both pigeons and humans (Elmore et al., 2012; Eng et al., 2005; Wright et al., 2010). Each object subtended a visual angle of $3.25^{\circ} \times 3.25^{\circ}$. All target objects were presented in the center of the screen. All objects presented in the probe array were located along the edge of two invisible concentric circles, with 6 locations on each circumference, creating 12 possible locations. The diameter of the inner circle subtended 72°, and the diameter of the outer circle subtended 111°. Additionally, the locations were rotationally offset from one another by 30°. Before presentation, the locations of probe objects were randomly jittered along both axes according to a Uniform distribution ranging from -0.2° to 0.2° .

2.1.4. Experimental control

All experimental events were controlled with custom software. Computer-controlled relay interfaces (Metrabyte, Freemont, CA) operated the food hopper and house light.



Fig. 1. (Left) Example trial progression for Experiments 1 (pigeons) and 2 (humans). (Right) Objects used in Experiments 1 (except Kaleidoscopes) and 2 (all objects). For the polygon and shape objects, the first four objects were used as the initial 4-object sets during pigeon training (Experiment 1) before expanding to all 8 objects.

2.1.5. Procedure

Two-alternative forced-choice training. The first phase of target search training began with a two-alternative forced-choice (2AFC) task using the 8 colored circles. Trials began with 15-s ITI during which the house light was turned on. After the house light was extinguished, a single colored circle was presented as the target in the center of the display. An observing response (FR1) to the circle produced a brief (500 ms) blank delay followed by a probe array containing two colored circles, located at random positions on the polar grid. Responses to the colored circle that matched the sample item produced 3-s access to mixed grain. Responses to the nonmatching colored circle (hereafter, the distractor) were followed by a correction procedure consisting of a 15-s dark timeout and repetition of the trial. The target and distractor were selected randomly without replacement from the 8-color set for each trial. Similarly, the locations of the probe array items were selected randomly without replacement from the set of 16 possible locations. The random selection of the probe item locations instead of three fixed locations distinguishes our training procedure from other matchto-sample procedures we (and many others) have used previously (e.g., Bodily et al., 2008). Subjects completed 96 trials within each session. The observing response to the target object was steadily increased over sessions to an FR10. After reaching a performance criterion of 85% correct the correction procedure was removed. Subjects were then required to have a single session with 80% correct without the correction procedure before moving to probe array expansion.

Probe array expansion. After reaching the performance criterion with the 2AFC task, the subjects completed sessions with equal mixtures (32 of each) of trials with probe array sizes 1-3. If accuracy dipped below 70%, the correction procedure was re-instated until accuracy went above 70%. The distractor items in the probe array were chosen randomly without replacement from the set of colors remaining after the target color was selected. Training continued until accuracy reached 80% (without correction procedure) on probe array size 3 trials, and then subjects were given sessions with equal mixtures (24 or each) of trials with probe array sizes of 1-4. In this manner, probe array size was increased via subsequent expansions so that eventually subjects were completing sessions with equal mixtures (16 of each) of trials with probe array sizes of 1-6 during the final condition of probe array training. When probe array size was expanded to 5, an equal mixture was not possible, and the extra trial was assigned as probe array size 1. Once performance reached 80% on probe array size 6 trials, subjects received a series of shape transfer tests.

Shape transfer series. After training with the color objects, subjects were transferred to a shape target search task via a series of 3 transfer tests, similar to the shape transfer tests used in our pigeon change detection study (Elmore et al., 2012). For each transfer test, eight of the probe array size 2 trials used shape objects rather than colored circles. The assignment of the 8 shape objects to target/nonmatching object was done pseudo-randomly so that each object was a target and non-matching probe item once per session. Additionally, no transfer trials were reversals of other transfer trials within the same session. For example, if one transfer trial used pentagon as target and heart as non-matching probe, then a trial with heart as target and pentagon as probe was never selected. Transfer tests were comprised of 6 sessions each and were conducted on consecutive days (18 days total). The correction procedure was never used during a transfer test.

In the first transfer test, the target item was still defined by its color, but rather than a circle, the shape was one of 7 novel shapes (see Fig. 1). Non-matching probe items had distinct colors but matched on shape. For example, if the target was a red pentagon, the non-matching probe item could be a yellow pentagon. The second transfer test used targets defined by both color and shape. For example, if the target was a red pentagon, the nonmatching probe item could be a yellow heart. The final transfer test used targets defined by shape only. For example, if the target was a red pentagon, the non-matching probe item could be a red heart.

Shape training. Subjects were given shape-only training immediately following the shape transfer series. Shape training began with probe array size 2 trials (2AFC), as in color training. To aid in the discrimination, only 4 shapes were used in training. The correction procedure was reinstated on the third day of shape training. After performance reached 80% the correction procedure was removed, and training continued until performance was at least 80% without correction procedure. Probe array size was expanded to four using the same criteria as color probe array expansion.

For J15313, after reaching the acquisition criterion for probe array size 2 trials, we integrated the remaining 4 untrained shape objects into the training set through 2 consecutive 6-session transfer tests. Each session in the first transfer tests contained 8 trials using only the 4 remaining shape objects. Each session in the second transfer test contained 16 transfer trials, evenly split into Target-Trained, Distractor-Untrained and Target-Untrained, Distractor-Trained types. We use the term untrained for the 4 shape objects used in these transfer tests to recognize that the objects are not, strictly speaking, novel, as they were used in the initial shape transfer test. After the two transfer tests with the untrained shape



Transfer Test Target Definition

Fig. 2. Baseline and transfer performance across shape and polygon transfer series for 2 pigeons. For the polygon transfer, the baseline trials used shape objects for target/distractor. For all other tests, the baseline trials used the color objects for target/distractor. Error bars represent 1 standard error of the mean.

objects, training continued with probe array size 2 trials using all 8 shape objects. Correction procedure was used beginning with session 3. From this point the probe array was expanded up to size 3, using the same criteria as with the color objects. Because J15313 could not meet the performance criteria to move to probe array size 4, we changed the object set back to the original 4 shape objects. Using just these 4 shapes, the subject was able to reach probe array size 4. After reaching probe array size 4, we alternated blocks of shape and color sessions in an attempt to increase performance. This training did not show any strong effect after extensive use (118 sessions).

Polygon transfer. After reaching 80% for a probe array size of 4 with the shape objects, subjects were given 6 sessions containing 8 transfer trials using polygon objects. All other procedural details were identical to the shape transfer tests.

Polygon training. Training with the polygon objects began with the use of 72 items and probe array sizes of 2 initially, but the set size was reduced to 4 polygons when J15313 did not reach high rates of accuracy in the task. Correction procedure was reinstated on the third session of polygon training. After subjects reached 85% accuracy, correction procedure was removed and a new criterion of 80% accuracy was set. After reaching this new criterion, subjects were given a single 6-session transfer test using 4 unfamiliar polygon objects from the 72-item set. After the transfer test, subjects received training sessions comprised of 8 polygon objects. After integrating the 8 polygon objects, training followed the procedure used for 8-color training, expanding the probe array up to size 4.

2.1.6. Results and discussion

Color training. Both subjects finished the initial 2-item, matchto-sample training within 20 sessions (J15313: 17 sessions; O13667: 19 sessions). J15313 reached the 80% performance criterion for probe array size 6 after 32 sessions, while O13667 needed 46 sessions.

Shape transfer. After training with the color objects, both subjects were given the shape transfer series. As can be seen in Fig. 2, both subjects performed well in the first two shape transfer tests, in which the target was still defined by color. Paired-sample *t*tests comparing mean baseline and transfer accuracy across the 6 transfer sessions for each transfer type (color, novel color and color+shape) and each subject confirmed this pattern [magnitude of all ts(5) < 1.97, ps > .1]. Once the target was defined solely by shape, however, neither subject's mean accuracy was above chance, as confirmed by per-subject one-sample *t*-tests [magnitude of both ts(5) < .83, ps > .44]. Additional analyses using the non-parametric Wilcoxon signed-rank test confirmed this pattern of transfer performance. These results suggest that both subjects were reliably using color to solve the discrimination and could not transfer this performance to a shape-only discrimination.

Shape training. J15313 required 41 sessions to reach the performance criterion of 80% without correction procedure with a probe array of 4 shape objects. O13667 required 55 sessions to reach the performance criterion of 80% without correction procedure with a probe array size of 4 objects. J15313 received an additional 166 sessions in an attempt to reach higher probe array sizes without success. Acquiring the shape discrimination was clearly more difficult than the color discrimination for the pigeons.

Polygon transfer and training. Results for the transfer to polygon objects are shown in Fig. 2. Similar to shape transfer, no evidence was found for transfer, as shown by per-subject one-sample *t*-tests testing transfer accuracy across the 6 sessions against chance performance [magnitude of both ts(5) < 1.40, ps > 0.221]. Non-parametric Wilcoxon signed-rank tests also confirmed that transfer performance was not significantly different than chance performance. Although this difference could be explained by restricted domain relational learning, we also note that the protracted shape training shows that discriminability may be an issue in locating targets defined solely by their shape.

Once J15313 proceeded to training with a set size of 8 polygons he needed 64 sessions to reach the 80% performance criterion for the 4-item probe arrays. It may be expected that successive acquisitions of new object sets may get progressively faster, as the subject learns more generally about the target search task. Because of the large difference between color and shape acquisition, however, it is clear that object discriminability remains an important factor. The next step is to compare search rates across the object types, to look for converging evidence about the role of complexity in pigeon VSTM.


Fig. 3. Target search slopes by object type for 2 pigeons. Points represent the mean of the last 6 sessions' median response times for each object type. Linear slope coefficients are listed by the corresponding line for statistically reliable (i.e., different from 0) slopes. Error bars represent 1 standard error of the mean.

Search rates across object types. Search rates were estimated for each subject by obtaining a least squares fit to the median response times for correct trials from each of the last 6 sessions of training with each object type. One subject (O13667) only had 4 sessions with all color probe array sizes; therefore only 4 sessions are used for this analysis. Fig. 3 shows the resultant search slopes for each subject by object type. Across subjects color objects were located the fastest, with neither slope being significantly different from 0 [013667: *F*(1, 22)=2.94, *p*=.100; [15313: *F*(1, 34 = 1.77, p = .192], although there is some indication of an increase in both subjects. Both subjects showed reliably positive slopes for shape target search [O13667: slope = 95.8, *F*(1, 22) = 56.72, *p* < .001; J15313: slope = 121.7, *F*(1, 22) = 19.52, *p* < .001], and J1531 showed a reliable positive slope for polygon target search [slope = 50.0, *F*(1, 22) = 7.02, p = .015]. Because of the visually apparent non-linearity in some of the search slopes, additional tests using a log transform on probe array size were conducted. We found reliable log-linear relationship between probe array size and color search times for O13667 [F(1, 22) = 4.60, p = .043], and between probe array size and polygon search times for J15313 [*F*(1, 22) = 9.51, *p* = .005]. No other log-linear relationships were reliable.

Because we did not counterbalance the order of object type training (or use a between subjects design as in Experiment 2), it is possible that the relative ease of polygon search as compared to shape search is due to the previous training with the shape stimuli. If this were the case, however, we would expect better transfer performance to polygons (Fig. 2) or much faster acquisition. Future work should consider how target search performance transfers across object type domains.

Comparison with change detection. Fig. 4 compares the average search slopes obtained from our target search task with the accuracies obtained in a series of transfer experiments (Elmore et al., 2012). The change detection data come from a series of transfer tests (using 2-item sample displays), rather than at the end of acquisition. The use of transfer data is important because our original question relates to reasons for failed transfer. If no relationship is found between target search rates and transfer performance, we could conclude that the lack of transfer is related mostly to the lack of relationship between search rates and transfer performance would suggest that object complexity may be moderating performance on transfer tests.

The relationship is clearly linear, and although there are only 3 object types, the target search slopes explain over 98% of the variance in change detection accuracy. The graph does reveal the surprising finding that both the polygon search rate and accuracy both fall in-between the values for color and shape. This finding is somewhat counter-intuitive as we initially designed the shape objects to be "simple," whereas the irregular polygons were thought to be more difficult. Converging evidence from both target search and change detection, however, suggest that pigeons perceived our shape objects as more complex than our polygon objects, within the set of objects they experienced. This discrepancy highlights the role of using target search to quantify visual object complexity, rather than relying on our own (species-specific) qualitative notion of object complexity.

As with the target search training, the single order of training in the change detection data (color, then shape, then polygon) leaves open the possibility that the better performance on polygon trials is explained by a general increase in performance. The lack of complete transfer from shapes to polygons (in both target search and change detection), however, suggests that pigeons had not learned a generalized change rule, but instead their relational domain was restricted to colors and the trained shapes.

3. Experiment 2

The goal of Experiment 2 was to collect target search data from humans in an analogous way as the pigeons in Experiment 1. We wanted to confirm the relationship found in previous human change detection studies (e.g., Alvarez and Cavanagh, 2004) would emerge using our task and a between-subjects manipulation. Finally, because of the difficulty of training pigeons to complete the task with highly complex objects, testing with humans allowed us to assess a wider range of object types across a large number of individuals.

3.1. Method

3.1.1. Participants

Participants were 48 students from Auburn University, who volunteered in return for course credit. Participants under 19 provided parental consent prior to the study, and all participants provided consent prior to beginning the experiment. The university's Institutional Review Board approved all experimental details.

3.1.2. Apparatus

Participants were seated approximately 30 cm away from a 17in. touchscreen LCD monitor ($1280 \times 1064, 60$ Hz). The touchscreen displayed all objects and recorded all responses made by the participants. The room in which participants were seated was unlit; the only light in the room came from the computer monitor. All



Fig. 4. Relationship between inverse percent correct in change detection and search rates from target search for pigeons (left panel) and humans (right panel).

sequences of the experiment were controlled and recorded using custom software.

3.1.3. Objects

Objects consisted of the same items from Experiment 1, with the addition of 8 kaleidoscope items (see Fig. 1). The visual angle for these objects was matched to the pigeons from Experiment 1.

3.1.4. Procedure

All participants were seated in front of the touchscreen. The experimenter read aloud task instructions and answered participant questions. Participants were randomly assigned to 1 of 4 visual object types (colors, shapes, irregular polygons, or kaleidoscopes). Participants completed 384 target search trials divided into 2 blocks of 192 trials, counterbalanced across participants for order. Each block contained 32 repetitions of 6 probe array sizes (1–6) pseudorandomly ordered.

As in Experiment 1, each trial began with the presentation of a target object, and progressed to a blank probe delay after an observing response. The observing response was set to a single touch for humans. A 500-ms probe delay separated the target and probe array. All objects in the probe array appeared at the same time and remained onscreen until a choice was made. Correct responses were followed by a 1000-ms tone. A 2000-ms intertrial interval (ITI) separated subsequent trials. Between the two blocks, participants were allowed to rest as needed.

3.1.5. Results and discussion

Target search slopes. Because participants were given explicit instructions, accuracies were uniformly high across probe array sizes and object types, and so we only consider response time in these analyses. Fig. 5 shows how search times varied across probe array size and object types. The most notable difference when compared with Fig. 3 is the lack of a strong distinction between the slopes for color and shape objects. Additionally, the polygon search times are slower than those for the shape objects.

We conducted a regression analysis on the median response times (correct trials only) for each subject across probe array size (probe array sizes 1 through 6; 1 parameter), object type (color, shape, polygon, and kaleidoscope; dummy-coded with color as

the baseline level providing 3 parameters), and interaction terms (formed by crossing the dummy-coded variables with the slope parameter; 3 parameters). Fitting this joint model provides direct tests of the slopes/intercepts of each object type against the baseline color model, rather than against 0 (if individual models were fit to each object type). Overall, our linear model of object type and probe array size accounted for 71% of the variance in median correct response times ($R_{adj}^2 = 0.707$). For the baseline color model, the best fitting intercept was 462.9 ms [t(280)=16.41, p<.001]and the slope was 24.29 ms/item [t(280)=3.353, p<.001]. For shape objects, the analysis showed a marginally reliable 78.3 ms increase in the intercept [t(280) = 1.96, p = .051], but no significant change in slope [t(280)=0.29, p=.776). For polygon objects, no change was found for intercept [t(280)=1.48, p=.140], but the slope increased reliably by 26.2ms [t(280)=2.56, p=.011]. Finally, for kaleidoscope objects, the intercept increased by a reliable 93.3ms [t(280)=2.34, p=.020] and the slope by a reliable 72.2 ms/item [t(280) = 7.25, p < .001]. A generalized least-squares (GLS) model using an auto-regressive moving-average [ARMA(1,2)] covariance structure showed comparable intercept/slope estimates and an identical pattern of parameter significance (including nonsignificance of the intercept increase for the shape object type, p = .096).

Comparison with change detection. As in Experiment 1, we next plotted the obtained target search rates with percent correct data from change detection (Elmore et al., 2013). The right panel of Fig. 4 shows a roughly linear relationship between inverse percent correct in change detection and the search rates from target search. The deviation from linearity is caused by the relative similarity in change detection accuracy for polygons and kaleidoscopes, but a large difference in the target search slopes. Human participants were not tested with shape objects (Elmore et al., 2013) so there is no corresponding shape data point. The target search data would suggest, however, that performance with the shapes should be nearly identical to the color-objects performance.

4. General discussion

Across two experiments, we obtained target search slopes for pigeons and humans and related these slopes to data from pigeon



Fig. 5. Target search slopes by object type for the human target search experiment. Points represent the average across each individual's median response time. Linear slope coefficients are listed by the corresponding line for statistically reliable (i.e., different from 0) slopes. Error bars represent 1 standard error of the mean.

and human change detection. Although the pigeons' experiences were not perfectly equated, the stability of the search functions shows the utility of the method. Pigeons' change detection performance was linearly related to the target search slopes for the range of objects tested. The function for the humans was broadly consistent with previously published human target search/change detection work (e.g., Alvarez and Cavanagh, 2004; Eng et al., 2005), and generally replicates the effect found in our pigeon task. Several comparisons between the functions suggest both quantitative and qualitative differences, within the general finding of a cross-species, complexity-limited VSTM. We note, however, that because only two pigeons were used to obtain the mean target search slopes, any generalization to the broader species must be appropriately tempered.

Across pigeons and humans, color objects produced the fastest search rates and highest change detection accuracies. Additionally, search rates increased and change detection accuracies decreased for polygon objects for both species. A difference arises for the shape objects, as pigeons experienced a nearly 6-fold increase in search rates, but humans showed no difference between color and shape target search. The validity of pigeons' difficulty with shape target search was confirmed by a similar difficulty in shape change detection. One reason for this difference may be that humans can verbally label the shape objects. Applying a label or category to objects facilitates memory performance in change detection (e.g., Olsson and Poom, 2005), and may provide a more efficient representation for target search. When compared to search for polygons (another type of object defined by its boundary) a dissociation emerges: polygon search rates were faster than shape for pigeons, but slower than shape search rates for humans, strengthening the labeling/categorization benefit for target search. The use of a verbal labeling strategy shows a qualitative difference in how humans and pigeons (and likely other nonhumans) solve discriminations more generally.

A few differences between the training of the pigeons and the humans should be mentioned. First, only pigeons had a correction procedure during the training. The correction procedure was necessary to ensure accurate responding by the pigeons in both tasks. During change detection transfer sessions (the sessions used for comparison with polygon and shape target search slopes), however, correction procedure was not in place, and is therefore unlikely to play a large part in our primary comparison between change detection performance and target search slopes. Secondly, humans used larger sample arrays during change detection and were tested with different object types within a session. This difference makes direct comparisons between human and pigeon change detection or target search difficult, but do not change our overall point that visual object complexity impacts memory performance in pigeons.

Despite the noted methodological differences, the general similarities between and within target search and change detection for each species are suggestive of a common "limiting factor" in how each species utilizes VSTM. Combining the current quantitative estimates of visual object complexity with change detection (Elmore et al., 2012) results in stronger tests of current evolutionary-based theories of VSTM. Gibson et al. (2011) have suggested that pigeon VSTM may be limited by a fixed capacity of around 2 items, similar to the roughly 4-item fixed capacity advanced for human VSTM (e.g., Luck and Vogel, 1997). Because the current pigeon change detection data come from sample arrays of only two items, however, all of the items should fit within their capacity. Under this model, then, performance declines in change detection not because of a reduction in the number (or resolution) of the items stored, but rather because of the increase in sample/probe array confusability that results from the increase in object complexity.

Alternatively, our results also fit within the continuous view of VSTM (e.g., Elmore et al., 2011; van den Berg et al., 2012). If VSTM is a continuous, flexibly allocated resource, the performance decrement across object types is explained by the difficulty of adequately encoding complex objects. A more complex object requires additional resources to be encoded with sufficient resolution to ensure a reasonable degree of accuracy. Differences between humans and pigeons with the least-complex objects (in the current study, colors) then arise from differences in the baseline noise levels of their respective VSTM systems.

Although the current results do not distinguish between discrete and continuous views of VSTM, they do provide more evidence of strong functional similarities between human and pigeon VSTM. Combined with recent work on nonhuman primate VSTM (Elmore et al., 2011; Heyselaar et al., 2011), it is becoming increasingly clear that a cross-species approach to VSTM can provide further insight and constraint on evolutionary theories of VSTM. The development of such a theory will be buoyed by continued emphasis on accurate estimation of cross-species functional relationships via parametric manipulation.

References

- Alvarez, G.A., Cavanagh, P., 2004. The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychol. Sci. 15, 106–111.
- Awh, E., Barton, B., Vogel, E.K., 2007. Visual working memory represents a fixed number of items regardless of complexity. Psychol. Sci. 18, 622–628.
- Bays, P.M., Husain, M., 2008. Dynamic shifts of limited working memory resources in human vision. Science 321, 851–854.
- Bodily, K.D., Katz, J.S., Wright, A.A., 2008. Matching-to-sample abstract-concept learning by pigeons. J. Exp. Psychol. Anim. Behav. Process. 34, 178–184.
- Elmore, L.C., Ma, W.J., Magnotti, J.F., Leising, K.J., Passaro, A.D., Katz, J.S., Wright, A.A., 2011. Visual short-term memory compared in rhesus monkeys and humans. Curr. Biol. 21, 975–979.

- Elmore, L.C., Magnotti, J.F., Katz, J.S., Wright, A.A., 2012. Change detection by rhesus monkeys (*Macaca mulatta*) and pigeons (*Columba livia*). J. Comp. Psychol. 126, 203–212.
- Elmore, L.C., Passaro, A., Wright, A.A., 2013. Change detection for study of object and location memory. Behav. Process. 93, 25–30.
- Eng, H.Y., Chen, D., Jiang, Y.V., 2005. Visual working memory for simple and complex visual object. Psychon. Bull. Rev. 12, 1127–1133.
- Gibson, B., Wasserman, E., Luck, S.J., 2011. Qualitative similarities in the visual short-term memory of pigeons and people. Psychon. Bull. Rev. 18, 979–984.
- Heyselaar, E., Johnston, K., Paré, M., 2011. A change detection approach to study visual working memory of the macaque monkey. J. Vision 11, 1–10.
- Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. Nature 390, 279–281.
- Olsson, H., Poom, L., 2005. Visual memory needs categories. Proc. Natl. Acad. Sci. U.S.A. 102, 8776–8778.
- van den Berg, R., Shin, H., Chou, W.-C., George, R., Ma, W.J., 2012. Variability in encoding precision accounts for visual short-term memory limitations. Proc. Natl. Acad. Sci. U.S.A. 109, 8780–8785.
- Wright, A.A., 2007. An experimental analysis of memory processing. J. Exp. Anal. Beh. 88, 405–433.
- Wright, A.A., Katz, J.S., Ma, W.J., 2012. How to be proactive about interference: lessons from animal memory. Psychol. Sci. 23, 453–458.
- Wright, A.A., Katz, J., Magnotti, J.F., Elmore, L.C., Babb, S., Alwin, S., 2010. Testing pigeon memory in a change detection task. Psychon. Bull. Rev. 17, 243–249.
- Zhang, W., Luck, S.J., 2008. Discrete fixed-resolution representations in visual working memory. Nature 453, 233–235.

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Review Remembering the past and planning for the future in rats^{\star}

Jonathon D. Crystal*

Department of Psychological & Brain Sciences, Indiana University, 1101 E 10th Street, Bloomington, IN 47405-7007, USA

ARTICLE INFO

Article history: Received 22 June 2012 Received in revised form 8 November 2012 Accepted 16 November 2012

Keywords: Animal model Comparative cognition Episodic memory Incidental encoding Prospective cognition Prospective memory

Contents

ABSTRACT

A growing body of research suggests that rats represent and remember specific earlier events from the past. An important criterion for validating a rodent model of episodic memory is to establish that the content of the representation is about a specific event in the past rather than vague information about remoteness. Recent evidence suggests that rats may also represent events that are anticipated to occur in the future. An important capacity afforded by a representation of the future is the ability to plan for the occurrence of a future event. However, relatively little is known about the content of represented future events and the cognitive mechanisms that may support planning. This article reviews evidence that rats remember specific earlier events from the past, represent events that are anticipated to occur in the future, and develops criteria for validating a rodent model of future planning. These criteria include representing a specific time in the future, the ability to temporarily disengage from a plan and reactivate the plan at an appropriate time in the future, and flexibility to deploy a plan in novel conditions.

© 2012 Elsevier B.V. All rights reserved.

1.	Introduction	39
2.	Remembering the past	40
	2.1. Do rats remember past episodes?	40
	2.2. Can memory be isolated to a specific past episode?	40
	2.3. Independent converging lines of evidence for episodic memory	42
3.	Planning for the future	45
	3.1. Prospective memory in the rat	45
	3.2. Evaluation of strengths and weaknesses	46
4.	Criteria to validate planning in rats	47
5.	Conclusions	47
	Acknowledgements	47
	References	47

1. Introduction

Memory enables information to be stored and retrieved after seconds to years and is essential for daily life. The loss of memory function is debilitating. Moreover, cognitive decline exerts significant societal costs. Consequently, even small improvements to retain cognitive function can have significant impacts on wellbeing, social engagement, and productivity. Episodic memory is memory for your own unique personal past experiences and the context in which those events occurred (Tulving, 1972); whereas episodic memories contain details about earlier events, semantic memory stores generic facts (Tulving, 1993). Episodic memory is one of the most vulnerable aspects of cognition. For example, episodic memory is profoundly impaired in patients suffering from Alzheimer's disease, and a decline in episodic memory is one of the earliest symptoms of Alzheimer's (Leube et al., 2008; Storandt, 2008). It is noteworthy that deficits in episodic memory in people afflicted with Alzheimer's are deficits in the *content* of episodic memory, not merely in reports of subjective experiences (Bäckman et al., 1999; Egerhazi et al., 2007; Le Moal et al., 1997; Liscic et al., 2007). Consequently, modeling the content of episodic memory in animals may facilitate development of therapeutic approaches for understanding and intervening in cognitive decline.

Prospective memory is the ability to remember to take some action in the future (McDaniel and Einstein, 2007). Because people





^{ightarrow} Jeffrey S. Katz was the acting editor for this paper.

^{*} Tel.: +1 812 856 2246; fax: +1 812 855 4691. *E-mail address:* jcrystal@indiana.edu

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.014

remember the past, they can remember to take actions in the future based on the past. Indeed, representing the future to simulate and predict possible future events depends on the same neural machinery that is used to remember the past (Schacter et al., 2007), including the medial prefrontal regions, posterior regions in the medial and lateral parietal cortex, the lateral temporal cortex, the medial temporal lobe, and hippocampus (Martin et al., 2011; Schacter et al., 2007). Integration of information from the past is used to construct simulations about future events (e.g., episodic simulation, planning, prediction, and remembering intentions) (Schacter et al., 2008). Consequently, it is not surprising that prospective memory is also impaired in Alzheimer's disease (Blanco-Campal et al., 2009; Driscoll et al., 2005; Jones et al., 2006; Schmitter-Edgecombe et al., 2009; Troyer and Murphy, 2007). Moreover, representations of past and future may share functional commonalities. For example, the temporal distance and elaboration of details regarding past and future events play similar roles in episodic memory and prospective cognition (Addis and Schacter, 2008; Crystal, 2012b; Roberts, 2012; Roberts and Feeney, 2009; Schacter and Addis, 2007), namely with remote events characterized by disparateness of details.

One benefit of studying cognition in animals is that it may provide insight into impairments in cognition observed in people. Developing insight into the origins of such impairments offers a tool to improve the effectiveness of treatments. Significant obstacles nonetheless impede the development of animal models of disordered cognition. Although there is a long history of studying learning and memory in animals, these types of cognitive processes may not match those observed clinically. For example, most preclinical models of Alzheimer's disease focus on general assessments of learning and memory, particularly spatial cognition, rather than on episodic memory. Thus, it is possible that drug-development programs may identify agents effective at the pre-clinical level that subsequently fail when translated to a clinical trial in people. Ultimately, the expansion of the suite of cognitive processes that may be modeled in animals may translate to improved therapies for debilitating memory impairments observed in humans (Crystal, 2012a).

2. Remembering the past

In the sections that follow, I examine three questions about remembering the past. (1) Do rats remember past episodes? (2) Can memory be isolated to a specific past episode, rather than general information about remoteness? (3) Can independent, converging lines of evidence be obtained that implicate the use of episodic memory?

2.1. Do rats remember past episodes?

One approach to demonstrating memory for a specific earlier event is to focus on what–where–when memory (Clayton et al., 2003; Clayton and Dickinson, 1998); that is, memory for what happened, where it took place, and when in time it occurred. Hence, we evaluated the hypothesis that rats have what–where–when memories while eliminating a number of non-episodic memory hypotheses.

Our approach (Zhou and Crystal, 2009, 2011) was to allow rats to encounter different flavors of food at various locations. We provided rats with daily information about the location of a preferred food type (chocolate) that replenished or did not replenish at its previously encountered location. Another flavor (regular chow) was available at all other locations but never replenished. The rats had the opportunity to learn that a preferred flavor would replenish at a recently presented location, but the replenishment was contingent on the time of day at which the flavor was initially encountered. If rats remember what-where-when they encountered the distinctive flavor, then they should revisit the distinctive location at a high rate when the distinctive flavor is about to replenish, but they should inhibit revisits on equivalent trials when the distinctive flavor is not about to replenish. By contrast, a rat without what-where-when memory would not be able to selectively revisit the location baited with the distinctive flavor more in the replenishment condition than in the non-replenishment condition.

In Zhou and Crystal (2009) study, rats' memory was assessed once per day, either in the morning or in the afternoon (see Fig. 1a). Chocolate replenished at a daily unique location at only one of these times of day (morning for some rats; afternoon for other rats). Another flavor (regular chow) was available at all other locations, but chow flavored locations never replenished. The interval between memory encoding (study phase) and memory assessment (test phases) was approximately 2 min. A rat with what-where-when memory could visit the chocolate location selectively on occasions when chocolate was about to replenish despite the fact that the location of chocolate varied randomly across days and the morning and afternoon sessions were presented in random order. Indeed, when the chocolate location was about to replenish, the rats revisited that location at a higher rate relative to equivalent trials in which chocolate did not replenish (Fig. 2a). Differential rates of revisiting chocolate-flavored locations was accomplished while rats accurately avoided revisits to depleted chow-flavored locations. These data are consistent with the hypothesis that rats used what-where-when memories to adjust revisit rates to the daily-unique chocolate location. Importantly, what-where-when memory in this study could not be based on the delay between study and test, which was constant in replenishment and non-replenishment conditions. Thus, several sources of vague information about remoteness (i.e., judging relative familiarity of the study items, judging how long ago the study occurred, or timing an interval between study and test) could not be responsible for selective revisits in the replenishment condition because the retention interval was constant in replenish and non-replenish conditions. This approach rules out important non-episodic memory solutions that have been difficult to control in earlier experiments (Babb and Crystal, 2006a; Clayton and Dickinson, 1998; Roberts et al., 2008).

2.2. Can memory be isolated to a specific past episode?

The *central hypothesis* in animal models of episodic memory is that, at the time of a memory assessment, the animal remembers a specific earlier event. According to this episodic memory hypothesis, at the time of memory assessment, the rats remembered the earlier study episode and adjusted revisits to chocolate at test accordingly. Although the study described above is consistent with the hypothesis that rats remember the earlier study episode, it is also possible that the rats solved the task by using a remaining piece of information about the remoteness of an earlier event, namely the light onset in the colony which was more remote in the afternoon than in the morning sessions. According to this non-episodic memory explanation, the rats may have been reactive at the time of test based on other available cues without remembering the study episode. Thus, Zhou and Crystal (2009) determined the type of timing mechanism used in what-where-when memory by testing the following two proposals. According to the circadian time-of-day hypothesis, the rats used a circadian signal (i.e., morning vs. afternoon) to adjust revisit rates at the daily-unique chocolate location; this view is consistent with the episodic-memory hypothesis that the rats remember the specific time of day at which the study episode occurred. Alternatively, according to the interval-timing hypothesis, the rats timed the interval from light onset in the colony to the morning and afternoon sessions. We employed a 6-h phase advance



Fig. 1. Schematic representation of experimental design of Zhou and Crystal's (2009) study. (a) Design of Experiment 1. First helpings (study phase; encoding) and second helpings (test phase; memory assessment) of food were presented either in the morning or afternoon, which was randomly selected for each session and counterbalanced across rats. Study and test phases show an example of the accessible arms, which were randomly selected for each rat in each session. Chocolate or chow flavored pellets were available at the distal end of four arms in the study phase (randomly selected). After a 2-min retention interval, the test phase provided chow-flavored pellets at locations that were previously blocked by closed doors. The figure shows chocolate replenished in the test phase conducted in the morning (7 a.m.) but not in the afternoon (1 p.m.), which occurred for a randomly selected half of the rats; these contingencies were reversed for the other rats (not shown). For each rat, one session was conducted per day. (b) Phase-shift design of Experiment 2. Light onset occurred at midnight, which was 6 hr earlier than in Experiment 1, and the session occurred in the morning. The horizontal lines emphasize the similarity of the 7-h gap between light onset and sessions in probe (solid) and training (dashed) conditions in Experiment 1. This design puts the predictions for time-of-day and how-long-ago cues in conflict; performance typical of the morning baseline is expected based on time of day whereas afternoon performance is expected based on how long ago. (c) Transfer-test design of Experiment 3. Study phases occurred at the same time of day as in Experiment 1. Test phases occurred at novel times of day (7 h later than usual). Therefore, early and late sessions had study times (but not test times) that corresponded to those in Experiment 1. The first two sessions in Experiment 3 were one replenishment and one non-replenishment condition, counterbalanced for order of presentation. An early or late session was randomly selected on subsequent days. More revisits to the chocolate location are expected in replenishment compared to non-replenishment conditions if the rats remembered the time of day at which the study episode occurred; revisit rates are expected to be equal in early and late sessions if the rats used the current time of day when the test phase occurred. Study and test phases were as in Experiment 1, except that they were separated by 7-h delays (shown by horizontal brackets). (d) Conflict-test design of Experiment 4. The study phase occurred at 1 p.m. and was followed by a test phase at 2 p.m. These times correspond to the time of day at which a late-session study phase and early-session test phase occurred in Experiment 3, which put predictions for time of day at study and time of day at test in conflict. If rats remembered the time of day at which the study episode occurred, they would be expected to behave as in its late-session, second-helpings baseline; alternatively, if the rats used the current time of day at test, they would be expected to behave as in its early-session, second-helpings baseline.

Reproduced with permission from Zhou and Crystal (2009). © 2009 National Academy of Sciences, U.S.A.

of the colony light cycle (i.e., light onset at midnight instead of 6 a.m.) and immediately tested the rats in the morning session (Zhou and Crystal, 2009). Because the magnitude of the phase advance was equal to the spacing between morning and afternoon sessions, an animal may treat the probe as an afternoon session (based on an interval mechanism) or as a morning session (based on a circadian mechanism). Importantly, adjusting the revisit strategy based on the passage of time since light onset can be done without remembering the time at which the study episode occurred, which makes the interval-timing proposal a non-episodic memory hypothesis. Because morning and afternoon sessions occurred 1 and 7 h, respectively, after light onset in the colony, a 6-h phase shift of light onset dissociates circadian time-of-day and interval-timing hypotheses. The lights in the colony were turned on 6 h early and the probe session was conducted at the usual time in the morning (see Fig. 1b). The rats treated the probe as a morning session (Fig. 2b), which is consistent with episodic memory because an endogenous circadian oscillator is not expected to adjust immediately to a phase shift (Takahashi et al., 2001). These data are significantly different from the predictions of the interval-timing hypothesis, according to which the rats would treat the probe as an afternoon session.

The studies outlined above suggest that rats can use time of day to judge when an event occurred in what-where-when memory. However, the central hypothesis about episodic memory is that the rats remember when the earlier study event occurred. By contrast, a non-episodic memory alternative hypothesis posits that the rats merely adjust revisit strategies based on the current time of day when a test occurs (Babb and Crystal, 2006b; Hampton and Schwartz, 2004; Nagshbandi et al., 2007). Thus, we next sought to determine if rats remember the time at which the earlier episode occurred (an episodic-memory hypothesis) or, alternatively, if they were merely selectively reactive at the different times of test. Importantly, reactivity at the time of test can occur without a memory of the earlier episode, making this a non-episodic memory alternative. Hence, we determined if it was the time of day at study or at test that was responsible for the different rates of revisiting the chocolate location. The rats had been trained using study and test times that were separated by a small (2 min) delay. Because this delay is too small for rats to discriminate based on a circadian oscillator (Pizzo and Crystal, 2004), we increased the delay between study and test to 7 h (see Fig. 1c), which is a value likely to be discriminated based on a circadian oscillator (Pizzo



Fig. 2. (a) Rats preferentially revisited the chocolate location when it was about to replenish in Experiment 1. The probability of a revisit to the chocolate location in the first four choices of a test phase is plotted for replenishment and non-replenishment conditions. (b) Rats used time of day, rather than information about remoteness, to adjust revisit rates in Experiment 2. The figure shows the difference between observed and baseline revisit rates. For the bar labeled interval, the baseline is the probability of revisiting chocolate in the afternoon. The significant elevation above baseline shown in the figure documents that the rats did not use remoteness or an interval mechanism. For the bar labeled time of day, the baseline is the probability of revisiting chocolate in the morning. The absence of a significant elevation above baseline is consistent with the use of time of day. The horizontal line corresponds to the baseline rate of revisiting the chocolate location in Experiment 1. Positive difference scores correspond to evidence against the hypothesis shown on the horizontal axis. (c and d) Rats preferentially revisited the replenishing chocolate location when the study, but not the test, time of day was familiar in Experiment 3. The probability of a revisit to the chocolate location in the first four choices of a test phase is shown for first replenishment and first non-replenishment sessions (c; initial) and for subsequent sessions (d; terminal). (e) Rats remembered the time of day at which the study episode occurred in Experiment 4. Rats treated the novel study-test sequence as a late-session test phase, documenting memory of the time of day at study rather than discriminating time of day at test. The figure shows the difference between observed and baseline revisit rates. For the bar labeled test time, the baseline was the probability of revisiting chocolate in the test phase of the early session in Experiment 3. The significant elevation above baseline documents that the rats did not use the time of day at test to adjust revisit rates. For the bar labeled study time, the baseline was the probability of revisiting chocolate in the test phase of the late session in Experiment 3. The absence of a significant elevation above baseline is consistent with memory of the time of day at study. The horizontal line corresponds to the baseline revisit rate to the chocolate location from Experiment 3 (terminal). Positive difference scores correspond to evidence against the hypothesis indicated on the horizontal axis. (a-e) Error bars indicate SEM. (a, c, and d) The probability expected by chance is 0.41. Repl = replenishment condition. Non-repl = non-replenishment condition. (a) *p < 0.001 difference between conditions. (b) *p < 0.04 different from baseline. (c and d) p < 0.04 and p < 0.001 difference between conditions. (e) p < 0.001 different from baseline. Reproduced with permission from Zhou and Crystal (2009). © 2009 National Academy of Sciences, U.S.A

and Crystal, 2006). Importantly, when we first introduced the long delay between study and test, the time of day at test was unfamiliar (approximately 7 h later than usual) whereas the time of day at study was familiar from earlier training. If the rats remembered the study episode, then they should continue to differentially revisit the chocolate locations when their memory was assessed at novel test times, using the familiar study time of day. Alternatively, if the rats were merely reactive to the time of day at test (i.e., without remembering the earlier study episode), then there is no basis for them to revisit chocolate locations at different rates in the morning and afternoon because the test times were unfamiliar; hence, the absence of episodic memory predicts equivalent revisit rates when tested at novel times of day. When tested with novel test times of day after familiar morning or afternoon study times of day, differential rates of revisiting occurred on the very first trial in the morning and afternoon (Fig. 2c and d, i.e., complete transfer). These data suggest that at the time of memory assessment, the rats remembered the time of day at which the study episode occurred.

Next, we obtained an additional line evidence for the same episodic-memory conclusion by putting episodic (study time) and non-episodic (test time) hypotheses into conflict. We used a novel combination of study and test times to determine if the rats remembered the study episode or were merely revisiting based on the current time of test. The 7-h delays between study and test phases produced a 1-h overlap between the two types of trials, which allowed us to start a trial with a *late* study phase and end the trial with an early test phase (see Fig. 1d). Again we sought to determine if the rats were adjusting revisit rates in the test phase based on the time of day at test (test-time hypothesis; the non-episodic memory proposal) or based on memory of the time of day at which the study phase occurred (study-time hypothesis; the episodic memory proposal). According to the test-time hypothesis, the rats should revisit at the baseline rate that was previously typical for that test time of day. Alternatively, according to the study-time hypothesis, the rats should revisit at the baseline rate that was previously typical for that study time of day. The rats adjusted chocolate revisits based on

the time of day at study rather than the time of day at test (Fig. 2e). These data suggest that rats remembered the study episode, and the time of day at which the study episode occurred, providing a second line of evidence that converges on the conclusion that rats remember when the earlier study episode occurred.

2.3. Independent converging lines of evidence for episodic memory

An important reason that episodic memory has been difficult to model in animals (and consequently controversial amongst researchers) is that behavioral training likely gives rise to welllearned expectations about the sequence of events. Thus, it is possible that animals may solve an episodic-memory test by using well-learned rules without remembering the episode at memory assessment; this possibility is a major threat to the validity of animal models of episodic memory. A fundamental aspect of episodic memory is that retrieval of information can occur when encoding is incidental and memory assessment is unexpected (Beran, 2012; Singer and Zentall, 2007a; Zentall, 2005, 2006, 2010; Zentall et al., 2001a, 2008; Zhou and Crystal, 2011; Zhou et al., 2012). Thus, we tested the hypothesis that rats can answer an unexpected question (via its behavior) after incidental encoding in a hippocampaldependent manner, consistent with the use of episodic memory (Zhou et al., 2012). Our approach builds on common features of episodic memory in the everyday life of people. Although events are not always known to be important when they occur, people can nonetheless report details about such events; in this situation, the memory assessment is unexpected, and the information is encoded incidentally (if it is encoded at all). For example, bystanders might observe a getaway car outside a crime scene. When the event occurs, it may not be obvious to observers that anything important has happened. However, during the subsequent investigation, important details about the getaway car may be obtained from eyewitnesses. This example highlights that such reports rely on memories for incidental aspects of the earlier episode.

Incidental encoding occurs when apparently unimportant information is stored, but it is not known at the time of encoding that the information may subsequently be quite useful. In other situations, information is explicitly encoded because the information is needed later. When information is encoded for use in an upcoming, expected test of retention, it is possible that success on the test is based on retrieval of a memory of the earlier episode. However, because the test is expected, it is also possible that the explicitly encoded information is used to generate a planned action: according to this view, at the time of the test, the remembered action can occur successfully without remembering the earlier episode. Thus, although explicit encoding and an expected test may yield successful performance, it is difficult to be certain that successful performance is based on a memory of the earlier episode. By contrast, when information is encoded incidentally, the nature of the subsequent memory test is not yet known, which prevents transforming the information into a specific action plan. Hence, if we observe accurate performance on an unexpected test after incidental encoding, it is likely that this performance is based on retrieval of an episodic memory. Most memory assessments in animals rely on explicit training (e.g., Clayton and Dickinson, 1998; Roberts et al., 2008; Zhou and Crystal, 2009, 2011). When an animal is trained to study some material and then repeatedly tested for retention, it is likely that studying gives rise to the expectation that the test will occur. Thus, whether the animal has episodic memory or not, it can perform accurately when information is encoded explicitly for an expected test.

Zentall (Singer and Zentall, 2007b; Zentall et al., 2001b, 2008) developed techniques to ask an animal an unexpected guestion after incidental encoding. Zentall et al. (2001a,b) trained pigeons in a symbolic matching task that was designed to determine if pigeons can answer the nonverbal question "Did you just peck or did you just refrain from pecking?" In one part of the experiment, the birds were trained to classify line orientations; the birds were trained on a symbolic matching task in which a line orientation (vertical vs. horizontal line) sample was followed by the requirement to peck or withhold pecking, followed by the selection of one of two colors (red vs. green). Note that the presentation of one line orientation signaled that a particular behavior (i.e., pecking or its absence) was required, which was then followed by the requirement to select one color to obtain reward. In another part of the experiment, the pigeons were provided with conditions that would elicit pecking or the absence of pecking, but without the requirement (and hence without an expectation) that a report about the pecking behavior would be required. In this part of the experiment, one color (e.g., yellow) was paired with food (which elicited pecking) and another color (e.g., blue) was presented but not paired with food (which elicited the absence of pecking). In the test, the sample stimuli that elicit pecking or the absence of pecking (i.e., yellow or blue) but that do not elicit the expectation of a question about pecking were presented. Next, the red and green comparison stimuli were presented, thereby unexpectedly providing the birds with the opportunity to report about their recent behavior (pecking vs. not pecking). When the pigeons were first asked the unexpected question, they reported accurately whether they had been pecking or not. In a further test, the birds were presented with a novel event that would elicit pecking (i.e., a new stimulus that occasioned generalized pecking) or a novel event that would elicit the absence of pecking (i.e., presentation of no stimulus on the test). Again the birds were unexpectedly asked whether they had recently pecked (i.e., by presentation of the red and green comparison stimuli), and they again accurately reported whether they had pecked or not. In further tests, Zentall and colleagues have controlled for residual proprioceptive cues that may be present when the unexpected question occurs (Singer and Zentall, 2007a) and shown that pigeons can also report about the location of their pecking response when unexpectedly asked (Zentall et al., 2008).

We recently developed techniques to test the hypothesis that rats can answer an unexpected question (via their behavior) after incidental encoding, and we used temporary inactivation of the hippocampus, an important processing center for episodic memory (Corkin, 2002; Eichenbaum, 2001; Ergorul and Eichenbaum, 2004; O'Brien and Sutherland, 2007; Tulving and Markowitsch, 1998; Vargha-Khadem, 1997), to test the hypothesis that answering an unexpected question requires episodic memory (Zhou et al., 2012). Importantly, at the time of encoding, it was not possible for the rat to know that the information would subsequently be needed (i.e., encoding was incidental) or that it would be requested (i.e., the test was unexpected). To determine if rats can answer an unexpected question after incidental encoding, we needed to train them to report about a recent event (in our case food vs. no food). Separately, we provided the rats with the opportunity to encounter food while foraging but where there was no expectation of being asked about the presence or absence of food. Next, we gave the rats the opportunity to incidentally encode the presence or absence of food while foraging and subsequently confronted them with the unexpected opportunity to report if they remembered encountering food or no food. Accordingly, we arranged for incidental encoding by using two types of tasks embedded within the same radial maze (Fig. 3). In one task, the rats foraged for food at multiple locations (five-arm radial maze task; Fig. 3a). In a second task, the rats learned the "reporting" skill (T-maze task; Fig. 3b) that would be used later in the unexpected question. In the T-maze task, rats were rewarded for selecting left/right turns after being presented with food or no-food, respectively. Because the animals received extensive training, the T-maze task involved explicit encoding for the purpose of answering an expected question. Thus, presentation of food or no food may have generated an action plan to turn left or right. Formally, an action plan based on semantic memory of a rule (e.g., if food \rightarrow turn left) may be formed when food/nofood occurs, but subsequently, the animal need only remember the to-be-performed response (left turn) without remembering the study episode. Thus, successful performance on the T-maze task does not specifically implicate the use of episodic memory. The purpose of the five-arm task was to provide the rats with an opportunity to search for food where there is no expectation of being asked about the presence of food. When foraging, the rat may encode the locations of food to avoid revisiting these locations or it may maintain a to-do-list of locations that have not yet been visited (Cook et al., 1985; Kesner, 1989). However, there was no expectation of being asked about the presence of food; thus, there was no reason for the rat to specifically plan to turn left/right. In this respect, the presence or absence of food during initial foraging is incidental to successfully obtaining additional food in foraging.

To generate incidental encoding and an unexpected question, rats began foraging for food and then were unexpectedly confronted with the opportunity to report whether they had recently encountered food or no food (Fig. 4a and b). Importantly, when unexpectedly confronted with the opportunity to report if it had recently encountered food or no food, to answer the unexpected question successfully, the rat would need to retrieve a memory of the earlier episode. Thus, a rat with episodic memory would be able to answer an unexpected question by retrieving a memory of the episode, despite the fact that the importance of the earlier encounter was not known at the time of encoding. By contrast, a rat without episodic memory would be unable to answer an unexpected question after incidental encoding; hence, the probability of left and right turns is expected to be equal in the absence of episodic memory. The dissociation of episodic memory is unique to unexpected questions after incidental encoding because both a rat with



Fig. 3. Schematic representation of experimental design of *training* from Zhou et al.'s (2012) study. (a) Five-arm task. Each rat was presented with study and test phases, separated by a brief retention interval (1 trial/day). An example of the accessible arms in the study phase and corresponding test phase is shown. Accessible arms were randomly selected for each rat on each session. Grey shading in the figure identifies arms used in the five-arm radial maze task. Doors to T-maze arms (shown in white) were closed. (b) T-maze task. Sample and choice phases were separated by a brief retention interval. In the sample phase, each rat was either given food (6 pellets) or no food (0 pellets). In the choice phase, each rat was rewarded with 6 pellets after turning left or right. Food and no-food samples led to reward in opposite sides of the T maze (counterbalanced across rats). Six trials were conducted per day with a random order of food and no-food samples. Doors to the five-arm radial maze were closed. (a and b) All arms of the actual maze were white.

Reproduced with permission from Zhou et al. (2012). © 2012 Elsevier Ltd.



Fig. 4. Schematic representation of experimental design of *probes* from Zhou et al.'s (2012) study. (a) Food and (b) no-food probes started with a study phase in the five-arm-radial-maze using arms situated 135°, 180° and 225° opposite to the sample arm. In the food probe, rats encountered one pellet at each of the three arms. In the no-food probe, rats visited these three arms but did not receive food pellets. Next, two choice arms from the T-maze were opened. (c) The rotation probe was identical to T-maze training (Fig. 3B), except the sample was presented in the arm 180° opposite to that used in T-maze training. (a–c) All arms of the actual maze were white. Reproduced with permission from Zhou et al. (2012). © 2012 Elsevier Ltd.



Fig. 5. (a) Rats answered unexpected questions after incidentally encoding the presence or absence of food. Baseline data come from the first daily T-maze trial in the terminal five days before probe testing. Food and no-food probes were each conducted once per rat. (b) Temporary inactivation of CA3 of the hippocampus before encoding impaired accuracy in answering an unexpected question relative to baseline but did not interfere with answering the expected question (rotation probe). Accuracy was selectively reduced by lidocaine in the unexpected probe relative to baseline and other probes. Baseline data come from the first daily T-maze trial in sessions before and after surgery. Each rat was tested once in each probe condition with the order counterbalanced according to a Latin Square design. Error bars represent 1 SEM. *p < 0.01 difference between the unexpected + lidocaine probe and baseline.

Reproduced with permission from Zhou et al. (2012). © 2012 Elsevier Ltd.

episodic memory and one without it would be able to answer an expected question after explicit encoding.

Rats answered unexpected questions (about the presence and absence of food; Fig. 4a and b) with a level of accuracy similar to that observed in training (Fig. 5a). We hypothesized that answering an unexpected question requires episodic memory. Thus, we next experimentally manipulated their ability to answer an unexpected question by temporarily inactivating the CA3 region of the hippocampus with bilateral infusions of lidocaine. To assess accuracy in answering an expected question, we used a control procedure (rotation probe, Fig. 4c) that preserved features of the T-maze task while equating other aspects of the no-food probe (e.g., extent of rotation).¹ Hippocampal inactivation selectively

eliminated the ability of rats to answer an unexpected question but did not affect performance on control conditions with an expected question (Figs. 4b, c and 5b). These experiments suggest that rats remember an earlier episode after incidental encoding based upon hippocampal-dependent episodic memory.

3. Planning for the future

Because representing the future to simulate and predict possible future events depends on the same neural machinery that is used to remember the past, it has been proposed that integration of information from the past is used to construct simulations about future events (Schacter et al., 2007, 2008). We recently provided evidence that rats remember to perform an intended future action, which suggests that rats posses at least a precursor to planning (Wilson and Crystal, 2012).

People "remember to remember." The hallmark of prospective memory is that, as the time to execute a remembered plan draws near, a deleterious effect on ongoing behavior occurs because greater attentional resources are diverted to the now activated prospective memory (Hicks et al., 2005; Kliegel et al., 2001; Marsh et al., 2006, 1998; Smith, 2003; Smith et al., 2007). According to this working model of prospective memory, when people form a prospective memory, they temporarily put the memory representation into an inactive state while engaging in other activities. Later, the representation is reactivated in the future. Ultimately, successful activation of the memory representation yields an action at an appropriate future time.² Prospective memory failures may occur when the memory representation fails to be reactivated at an appropriate time.

3.1. Prospective memory in the rat

We recently developed an animal model of prospective memory (Wilson and Crystal, 2012). The basic insight is that prospective memory produces a selective deficit in performance at the time when anticipation of a future event is greatest. To provide an ongoing activity, rats were trained in a temporal bisection task for 90 min

Prospective memory is distinguished from prospective coding, the latter of which has been more extensively studied in animals (e.g., Roitblat, 1980; Zentall, 2005). For example, in a delayed symbolic matching to sample experiment, a pigeon is presented with a sample (e.g., red or green) which predicts the subsequently rewarded choice (e.g., vertical vs. horizontal lines). An animal may solve this matching task by maintaining a representation of the studied sample (a retrospective code). Alternatively, an animal may solve this task by using a transformation rule (e.g., if red then choose vertical) to translate a code for the presented sample into a code for the forthcoming correct choice and from that point onward maintain a representation of the to-be-selected choice (a prospective code). Note that in both retrospective and prospective coding, the animal is hypothesized to maintain a memory code throughout the retention delay. However, the subsequent action (a correct response after the delay) is not fundamentally different whether the code was retrospective or prospective, except for the use of the transformation rule. By contrast, prospective memory is proposed to involve activating a representation, inactivating the representation, and then subsequently *reactivating* the representation at a later time. Inactivation and reactivation processes are not involved in prospective coding, unlike prospective memory

¹ We argued that the rotation probe does not require episodic memory for the following reasons. First, after the study phase, there is nothing unexpected about the test. Second, the study phase is identical to training (despite using a different start location) for a rat that relies on a response-mediated strategy; our rats had received extensive training in the T-maze task prior to probe testing, by which point they would likely rely on a striatal-response system (De Leonibus et al., 2011; Packard, 1999; Packard and McGaugh, 1996; Yin and Knowlton, 2004). Thus, the ability to solve the rotation probe was not expected to require an intact

hippocampus because well-trained habits have previously been suggested to be striatal-dependent (Packard, 1999; Packard and McGaugh, 1996). Moreover, it is unlikely that rats expected "unexpected" questions for the following reasons. First, the rats had received many 5-arm study phases that were not followed by an assessment of food/no-food. Second, right- and left-turn responses in the 5-arm task had an equivalent history of reinforcement, which was produced by the random selection of arm baiting in the study phase of the 5-arm task. Third, hippocampal inactivation eliminated the ability to answer the unexpected but not the expected question. By contrast, the striatum, which underlies habit learning, may mediate the ability of rats to answer the expected question (De Leonibus et al., 2011; Packard, 1999; Packard and McGaugh, 1996; Yin and Knowlton, 2004). Finally, although radial maze tasks typically use baited locations, eating is incidental to efficient navigation (Timberlake and White, 1990).



Fig. 6. Performance in an ongoing task was selectively impaired near the time of an anticipated future event. The probability of judging an interval as long (a) increased as a function of the interval duration, as expected. Performance, as measured by the slope of the probability function (b) declined immediately before the end of the daily session in the meal group (a) but not in the no-meal group (not shown). Importantly, the interaction between early vs. late time points and duration was significant for the meal group (a, p < 0.001) but not for the no-meal group (p = 0.1), and these group differences were significant as documented by the three-way interaction (p < 0.009). Similarly, the slope of the psychophysical function was smaller (i.e., poor performance) at the late relative to early time points (p = 0.009) in the meal group but not in the no-meal group (p=0.3), and these group differences were significant as documented by the arrival of the meal, as documented by the increase in food-trough responses before the meal whereas the increase in food-trough responses was absent in the no-meal group (c). These data document a selective impairment in performance in an ongoing task near the time of an anticipated future event (but not at other times). (a-c) Error bars indicate SEM. Reproduced from Wilson and Crystal (2012). © 2011 Springer-Verlaz.

per day. In bisection trials a 2- or 8-s signal was presented, and a small reward was delivered if the rat pressed the correct lever to classify the signal as short or long. To provide an anticipated future event, rats in the meal group earned an 8-g meal when the bisection task ended, whereas other rats in the no-meal group received no additional food. The meal was earned by interrupting a photobeam located inside a food trough, but photobeam breaks were only effective 90 min after the start of the bisection task. Rats in the meal group may remember to collect the meal, whereas rats in the no-meal group did not learn to remember an additional action beyond the bisection task. If rats have prospective memory, then the meal group should exhibit a negative side effect on ongoing task performance at a late time point (when the representation is most likely activated). If rats do not have prospective memory, then any change in performance from early to late time points should be equivalent for both meal and no-meal groups.

Wilson and Crystal (2012) showed that performance in the ongoing task declined near the meal time in the meal group but not in the no-meal group, consistent with prospective memory. Temporal sensitivity (i.e., the steepness of the psychophysical functions) declined near the meal time in the meal group but not in the no-meal group, as predicted by the prospective-memory hypothesis. Performance in the ongoing task was examined at early and late time points (Fig. 6). Temporal sensitivity decreased from early to late time points for the meal group but not for the no-meal group, as predicted by prospective memory. Because ongoing task performance is relatively constant throughout the session when a representation of a meal is absent, our findings suggest that the approaching meal produced the observed performance decline in the meal group. Food-trough visits increased as a function of time in the meal group but not in the no-meal group (Fig. 6), which suggests that the meal group anticipated the arrival of the meal, as expected.

We proposed that prospective memory produced the decline in ongoing task performance as the meal approached (Wilson and Crystal, 2012). According to this view, rats formed a representation of the meal but inactivated it at the early time point, when the meal was distant. As the expectation of the meal grew, more attentional resources were recruited to maintain the representation of the forthcoming meal, which impaired ongoing task performance. To support the representational account, we ruled out four non-representational hypotheses (attentional limit, response competition, contrast, and fatigue). A non-representational hypothesis is unlikely to explain our data because (1) an attentional limit imposed by judging intervals and anticipating the meal predicts impaired bisection performance throughout the entire pre-meal window, which is contrary to our data; (2) response competition (trough visits cause a decline in bisection performance) predicts a negative correlation between trough visits and bisection performance, which is contrary to our data; (3) contrast (diminished reward value in the bisection task in anticipation of higher reward value during the meal) predicts a decline in motivation to lever press measured by latencies, which is contrary to our data; and (4) fatigue (more behavioral output in meal than no-meal group) also predicts an increase in latencies to lever press, which is contrary to our data (for details see Wilson and Crystal, 2012). Another possible explanation for disruption is that a forthcoming meal may produce arousal, although this is unlikely for the following reason. Although arousal is associated with access to food (Bizo and White, 1994; Killeen et al., 1978), timing theories based on arousal propose that the rate of previously presented food sets a single level of arousal (Killeen and Fetterman, 1988). Thus, if arousal produces deleterious effects, the impact should be equivalent at early and late timepoints (during which the rate of reward did not change). Nevertheless, a direct test of the hypothesis that arousal may change dynamically will require assessments of deleterious effects (before, during, and after) experimental manipulations of arousal.

3.2. Evaluation of strengths and weaknesses

Our initial attempt to document prospective memory in rats can be evaluated with respect to its strengths and weaknesses.

The major strength of the model is that it provides a method for evaluating the existence of a representation of a future event that would otherwise be behaviorally silent. The observation that anticipating the arrival of the meal produces a deleterious effect on ongoing behavior suggests that rats form a prospective memory of the future meal. Because the model permits study of a representation of a future event, it may be possible to use the model to study the biological basis of human memory disorders. For example, failures of prospective memory (i.e., forgetting to act on an intention at an appropriate time in the future) are a common feature of aging (Aberle et al., 2010; Craik, 1986; d'Ydewalle et al., 2001; Driscoll et al., 2005; Henry et al., 2004) and negatively impacts both health (e.g., forgetting to take medications (Woods et al., 2009)) and independence (Mateer et al., 1987) (e.g., forgetting to lock one's home, turn off the stove, etc.). Prospective memory is impaired in a number of clinical populations, including patients with mild cognitive impairment (Schmitter-Edgecombe et al., 2009; Troyer and Murphy, 2007), Alzheimer's disease (Blanco-Campal et al., 2009; Jones et al., 2006; Troyer and Murphy, 2007), Parkinson disease (Foster et al., 2009; Raskin et al., 2011), traumatic brain injury (Henry et al., 2007; Mateer et al., 1987; McCauley et al., 2009), and HIV infection (Carey et al., 2006; Woods et al., 2009, 2006). Thus, the animal model of prospective memory may be a valuable tool to explore the biological bases of prospective memory disorders.

There are potential limitations of the model, especially given that not many predictions of the model have been tested. Rats show a cognitive side effect of prospective memory-representation of a future event (Addis and Schacter, 2008; Martin et al., 2011; Roberts and Feeney, 2009; Schacter and Addis, 2007; Schacter et al., 2007, 2008). However, it is not known if rats represent a specific time point in the future. By contrast, human prospective cognition includes the ability to specify a time point for both episodic memories about the past and planning for a specific time in the future - rather than merely general knowledge about remoteness -(Crystal, 2012b; Roberts, 2012; Roberts and Feeney, 2009; Schacter and Addis, 2007). The specificity in time is particularly important for time-based prospective memory in which people reactivate a memory representation at an appropriate point in time; by contrast, temporal specificity may be less important in event-based prospective memory in which people reactivate a memory representation when a particular event occurs (McDaniel and Einstein, 2007).

Prospective memory in people involves forming a plan, *inactivating* the representation, and then reactivating it in the future. It is not known if rat prospective memory can withstand a disengagement from the plan (i.e., inactivation), or alternatively if continual engagement is required.

People show a high degree of flexibility in planning. Although rats show a cognitive side effect of planning to obtain a meal in the future, it is not known if this ability is limited to conditions in which it is extensively trained, which would be the case if prospective memory in rats is limited to a learned fixed sequence of actions based on reflexive mechanisms.

4. Criteria to validate planning in rats

The review of limitations of the initial prospective memory model, suggests steps that are needed to test the model. In the section below, three criteria for validating an animal model of prospective memory are outlined.

First, the techniques that we have used to establish that rats remember a specific point in the past (reviewed above) can be adapted to test the hypothesis that rats *represent a specific point in time in the future* (Crystal, 2012b). Second, tests are required to determine if rats can temporarily disengage from a plan and subsequently reactivate the plan at an appropriate time in the future. Distractor tasks can be used to require the rat to disengage from the plan. Beran et al. (2012) have recently provided an excellent demonstration of prospective memory in a language trained chimpanzee, which included a clearer demonstration of disengagement (see also Evans and Beran, 2012). Third, tests are required to determine if rats exhibit some degree of flexibility to deploy a learned plan in a novel context.

5. Conclusions

A substantial body of research strongly suggests that rats use episodic memory to remember the content of specific earlier events. The content of these representations include the time of occurrence of the earlier event, where it occurred in space, and what flavor of food was encountered (for evidence of flavor specificity see Babb and Crystal, 2006a). Moreover, rats use episodic memory to retrieve information about events that were not known to be important at the time of encoding when they are unexpectedly asked to report about this information. Our approach provides an animal model of prospective memory (Wilson and Crystal, 2012), yet relatively little is known about the content of represented future events and the basic cognitive mechanisms that may support planning in rats. This review has highlighted some directions to explore, including the temporal specificity of prospective cognition, the ability to disengage and subsequently reactivate a representation of a future event, and the range of flexibility or creativity within prospective cognition.

A multi-method approach is needed to fully explore the elements of prospective cognition in rats. It is possible that rats have some aspects of prospective cognition, but in some significant ways it may be limited relative to prospective cognition in humans or other animals. The use of multiple approaches is likely to provide a more complete picture of the representations used in prospective cognition. Maintaining a representation of a future event is a prerequisite for planning yet fully developed planning may not be implicated. For example, other studies of planning (Cheke and Clayton, 2012; Correia et al., 2007; Mulcahy and Call, 2006; Naqshbandi and Roberts, 2006; Raby et al., 2007) document that some animals take action now for a future need that is dissociated from their current motivational needs (Suddendorf and Corballis, 1997, 2007). According to this mental time travel approach, an animal forms a representation in which it envisions itself in a future scenario. By contrast, in our approach, rats were food restricted and participated in two tasks that both provided food. Thus, our approach clearly did not seek to dissociate motivational states. Moreover, no evidence for planning was obtained in experiments that dissociated motivational states in rats (Nagshbandi and Roberts, 2006). Thus, it is possible that rats exhibit a precursor to planning only in a limited sense, and they may not be capable of more robust planning; alternatively, a multi-method approach with refinements in techniques may reveal more robust planning in future research. Although significant progress has been made using the mental-time-travel framework, it has recently been argued that future-oriented cognition should also be evaluated outside this framework (Crystal, 2012b; Raby and Clayton, 2009; Zentall, 2006, 2010). One advantage of our approach to model prospective memory in rats outside the mental-time-travel framework is that it may provide insight into the evolution of planning to act in the future across a wide array of species by focusing on deleterious side-effects of a prospective memory representation. Preserving the ability to evaluate prospective cognition in a wide range of species will be valuable for future research that seeks to exploit rodent models of human diseases with impaired cognition.

Acknowledgements

This article is in honor of the contributions of Tony Wright to the study of comparative cognition. Supported by National Institute of Mental Health grant R01MH080052.

References

- Aberle, I., Rendell, P.G., Rose, N.S., McDaniel, M.A., Kliegel, M., 2010. The age prospective memory paradox: Young adults may not give their best outside of the lab. Dev. Psychol. 46, 1444–1453.
- Addis, D.R., Schacter, D.L., 2008. Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. Hippocampus 18, 227–237.
- Babb, S.J., Crystal, J.D., 2006a. Discrimination of what, when, and where is not based on time of day. Learn. Behav. 34, 124–130.
- Babb, S.J., Crystal, J.D., 2006b. Episodic-like memory in the rat. Curr. Biol. 16, 1317–1321.

Bäckman, L., Andersson, J.L., Nyberg, L., Winblad, B., Nordberg, A., Almkvist, O., 1999. Brain regions associated with episodic retrieval in normal aging and Alzheimer's disease. Neurology 52, 1861–1870.

- Beran, M.J., 2012. Animal memory: rats can answer unexpected questions about past events. Curr. Biol. 22, R491–R493.
- Beran, M.J., Perdue, B.M., Bramlett, J.L., Menzel, C.R., Evans, T.A., 2012. Prospective memory in a language-trained chimpanzee (*Pan troglodytes*). Learn. Motiv. 43, 192–199.

Bizo, L.A., White, K.G., 1994. The behavioral theory of timing: reinforcer rate determines pacemaker rate. J. Exp. Anal. Behav. 61, 19.

- Blanco-Campal, A., Coen, R.F., Lawlor, B.A., Walsh, J.B., Burke, T.E., 2009. Detection of prospective memory deficits in mild cognitive impairment of suspected Alzheimer's disease etiology using a novel event-based prospective memory task. J. Int. Neuropsychol. Soc. 15, 154–159.
- Carey, C.L., Paul Woods, S., Rippeth, J.D., Heaton, R.K., Grant, I., the HIV Neurobehavioral Research Center (HNRC) Group, 2006. Prospective memory in HIV-1 infection. J. Clin. Exp. Neuropsychol. 28, 536–548.
- Cheke, L.G., Clayton, N.S., 2012. Eurasian jays (*Garrulus glandarius*) overcome their current desires to anticipate two distinct future needs and plan for them appropriately. Biol. Lett. 8, 171–175.
- Clayton, N.S., Bussey, T.J., Dickinson, A., 2003. Can animals recall the past and plan for the future? Nat. Rev. Neurosci. 4, 685–691.
- Clayton, N.S., Dickinson, A., 1998. Episodic-like memory during cache recovery by scrub jays. Nature 395, 272–274.
- Cook, R.G., Brown, M.F., Riley, D.A., 1985. Flexible memory processing by rats: use of prospective and retrospective information in the radial maze. J. Exp. Psychol. Anim. Behav. Process. 11, 453–469.
- Corkin, S., 2002. What's new with the amnesic patient H.M.? Nat. Rev. Neurosci. 3, 153–160.
- Correia, S.P.C., Dickinson, A., Clayton, N.S., 2007. Western scrub-jays anticipate future needs independently of their current motivational state. Curr. Biol. 17, 856–861.
- Craik, F.I.M., 1986. A functional account of age differences in memory. In: Hagendorf, F.K.H. (Ed.), Human memory and Cognitive Capabilities: Mechanisms and Performances. Elsevier, North Holland.
- Crystal, J.D., 2012a. Animal models of human cognition. In: Vonk, J., Shackelford, T. (Eds.), Oxford Handbook of Comparative Evolutionary Psychology. Oxford University Press, Oxford.
- Crystal, J.D., 2012b. Prospective cognition in rats. Learn. Motiv. 43, 181-191.
- d'Ydewalle, G., Bouckaert, D., Brunfaut, E., 2001. Age-related differences and complexity of ongoing activities in time- and event-based prospective memory. Am. J. Psychol. 114, 411–423.
- De Leonibus, E., Costantini, V.J.A., Massaro, A., Mandolesi, G., Vanni, V., Luvisetto, S., Pavone, F., Oliverio, A., Mele, A., 2011. Cognitive and neural determinants of response strategy in the dual-solution plus-maze task. Learning & Memory 18, 241–244.
- Driscoll, I., McDaniel, M.A., Guynn, M.J., 2005. Apolipoprotein E and prospective memory in normally aging adults. Neuropsychology 19, 28–34.
- Egerhazi, A., Berecz, R., Bartok, E., Degrell, I., 2007. Automated Neuropsychological Test Battery (CANTAB) in mild cognitive impairment and in Alzheimer's disease. Prog. Neuropsychopharmacol. Biol. Psychiatry 31, 746–751.
- Eichenbaum, H., 2001. The hippocampus and declarative memory: cognitive mechanisms and neural codes. Behav. Brain Res. 127, 199–207.
- Ergorul, C., Eichenbaum, H., 2004. The hippocampus and memory for 'what', 'where', and 'when'. Learn. Memory 11, 397–405.
- Evans, T.A., Beran, M.J., 2012. Monkeys exhibit prospective memory in a computerized task. Cognition 125, 131–140.
- Foster, E.R., McDaniel, M.A., Repovš, G., Hershey, T., 2009. Prospective memory in Parkinson disease across laboratory and self-reported everyday performance. Neuropsychology 23, 347–358.
- Hampton, R.R., Schwartz, B., 2004. Episodic memory in nonhumans: what, and where, is when? Curr. Opin. Neurobiol. 14, 192–197.
- Henry, J.D., MacLeod, M.S., Phillips, L.H., Crawford, J.R., 2004. A meta-analytic review of prospective memory and aging. Psychol. Aging 19, 27–39.
- Henry, J.D., Phillips, L.H., Crawford, J.R., Kliegel, M., Theodorou, G., Summers, F., 2007. Traumatic brain injury and prospective memory: influence of task complexity. J. Clin. Exp. Neuropsychol. 29, 457–466.
- Hicks, J.L., Marsh, R.L., Cook, G.I., 2005. Task interference in time-based, event-based, and dual intention prospective memory conditions. Journal of Memory and Language 53, 430–444.
- Jones, S., Livner, Å., Bäckman, L., 2006. Patterns of prospective and retrospective memory impairment in preclinical Alzheimer's disease. Neuropsychology 20, 144–152.
- Kesner, R.P., 1989. Retrospective and prospective coding of information: role of the medial prefrontal cortex. Exp. Brain Res. 74, 163–167.
- Killeen, P.R., Fetterman, J.G., 1988. A behavioral theory of timing. Psychol. Rev., 95. Killeen, P.R., Hanson, S.J., Osborne, S.R., 1978. Arousal: Its genesis and manifestation as response rate. Psychol. Rev. 85, 571–581.
- Kliegel, M., Martin, M., McDaniel, M.A., Einstein, G.O., 2001. Varying the importance of a prospective memory task: differential effects across time – and event-based prospective memory. Memory 9, 1–11.
- Le Moal, S., Reymann, J.M., Thomas, V., Cattenoz, C., Lieury, A., Allain, H., 1997. Effect of normal aging and of Alzheimer's disease on episodic memory. Dement. Geriatr. Cogn. Disord. 8, 281–287.

- Leube, D.T., Weis, S., Freymann, K., Erb, M., Jessen, F., Heun, R., Grodd, W., Kircher, T.T., 2008. Neural correlates of verbal episodic memory in patients with MCI and Alzheimer's disease—a VBM study. Int. J. Geriatr. Psychiatry 23, 1114–1118.
- Liscic, R.M., Storandt, M., Cairns, N.J., Morris, J.C., 2007. Clinical and psychometric distinction of frontotemporal and Alzheimer dementias. Arch. Neurol. 64, 535–540.
- Marsh, R.L., Hicks, J.L., Cook, G.I., 2006. Task interference from prospective memories covaries with contextual associations of fulfilling them. Memory Cogn. 34, 1037–1045.
- Marsh, R.L., Hicks, J.L., Landau, J.D., 1998. An investigation of everyday prospective memory. Memory Cogn. 26, 633–643.
- Martin, V.C., Schacter, D.L., Corballis, M.C., Addis, D.R., 2011. A role for the hippocampus in encoding simulations of future events. Proc. Natl. Acad. Sci. 108, 13858–21386.
- Mateer, C.A., Sohlberg, M.M., Crinean, J., 1987. Focus on clinical research: perceptions of memory function in individuals with closed-head injury. J. Head Trauma Rehab. 2, 74–84.
- McCauley, S.R., McDaniel, M.A., Pedroza, C., Chapman, S.B., Levin, H.S., 2009. Incentive effects on event-based prospective memory performance in children and adolescents with traumatic brain injury. Neuropsychology 23, 201–209.
- McDaniel, M.A., Einstein, G.O., 2007. Prospective Memory: An Overview and Synthesis of an Emerging Field. Sage Publications, Thousand Oaks, CA.
- Mulcahy, N.J., Call, J., 2006. Apes save tools for future use. Science 312, 1038-1040.
- Naqshbandi, M., Feeney, M.C., McKenzie, T.L.B., Roberts, W.A., 2007. Testing for episodic-like memory in rats in the absence of time of day cues: replication of Babb and Crystal. Behav. Process. 74, 217–225.
- Naqshbandi, M., Roberts, W.A., 2006. Anticipation of future events in squirrel monkeys (*Saimiri sciureus*) and rats (*Rattus norvegicus*): tests of the Bischof-Kohler hypothesis. J. Comp. Psychol. 120, 345–357.
- O'Brien, J., Sutherland, R.J., 2007. Evidence for episodic memory in a pavlovian conditioning procedure in rats. Hippocampus 17, 1149–1152.
- Packard, M.G., 1999. Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. Proc. Natl. Acad. Sci. U.S.A. 96, 12881–21288.
- Packard, M.G., McGaugh, J.L., 1996. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol. Learn. Mem. 65, 65–72.
- Pizzo, M.J., Crystal, J.D., 2004. Time-place learning in the eight-arm radial maze. Learn. Behav. 32, 240–255.
- Pizzo, M.J., Crystal, J.D., 2006. The influence of temporal spacing on time-place discrimination. Learn. Behav. 34, 131–143.
- Raby, C.R., Alexis, D.M., Dickinson, A., Clayton, N.S., 2007. Planning for the future by western scrub-jays. Nature 445, 919–921.
- Raby, C.R., Clayton, N.S., 2009. Prospective cognition in animals. Behav. Proc. 80, 314–324.
- Raskin, S.A., Woods, S.P., Poquette, A.J., McTaggart, A.B., Sethna, J., Williams, R.C., Tröster, A.I., 2011. A differential deficit in time- versus event-based prospective memory in Parkinson's disease. Neuropsychology 25, 201–209.
- Roberts, W.A., 2012. Evidence for future cognition in animals. Learn. Motiv. 43, 169–180.
- Roberts, W.A., Feeney, M.C., 2009. The comparative study of mental time travel. Trends Cogn. Sci. 13, 271–277.
- Roberts, W.A., Feeney, M.C., MacPherson, K., Petter, M., McMillan, N., Musolino, E., 2008, Episodic-like memory in rats: is it based on when or how long ago? Science 320, 113–115.
- Roitblat, H., 1980. Codes and coding processes in pigeon short-term memory. Learn. Behav. 8, 341–351.
- Schacter, D.L., Addis, D.R., 2007. The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos. Trans. Roy. Soc. B: Biol. Sci. 362, 773–786.
- Schacter, D.L., Addis, D.R., Buckner, R.L., 2007. Remembering the past to imagine the future: the prospective brain. Nat. Rev. Neurosci. 8, 657–661.
- Schacter, D.L., Addis, D.R., Buckner, R.L., 2008. Episodic simulation of future events. Ann. N.Y. Acad. Sci. 1124, 39–60.
- Schmitter-Edgecombe, M., Woo, E., Greeley, D.R., 2009. Characterizing multiple memory deficits and their relation to everyday functioning in individuals with mild cognitive impairment. Neuropsychology 23, 168–177.
- Singer, R., Zentall, T., 2007a. Pigeons learn to answer the question where did you just peck? And can report peck location when unexpectedly asked. Learn. Behav. 35, 184–189.
- Singer, R.A., Zentall, T.R., 2007b. Pigeons learn to answer the question 'where did you just peck?' and can report peck location when unexpectedly asked. Learn. Behav. 35, 184–189.
- Smith, R.E., 2003. The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. J. Exp. Psychol.—Learn. Mem. Cogn. 29, 347–361.
- Smith, R.E., Hunt, R.R., McVay, J.C., McConnell, M.D., 2007. The cost of event-based prospective memory: salient target events. J. Exp. Psychol.—Learn. Mem. Cogn. 33, 734–746.
- Storandt, M., 2008. Cognitive deficits in the early stages of Alzheimer's disease. Curr. Dir. Psychol. Sci. 17, 198–202.
- Suddendorf, T., Corballis, M.C., 1997. Mental time travel and the evolution of the human mind. Genet. Soc. Gen. Psychol. Monogr. 123, 133–167.
- Suddendorf, T., Corballis, M.C., 2007. The evolution of foresight: what is mental time travel, and is it unique to humans? Behav. Brain Sci. 30, 299–313.

- Takahashi, J.S., Turek, F.W., Moore, R.Y., 2001. Handbook of Behavioral Neurobiology: Circadian Clocks. Plenum, New York.
- Timberlake, W., White, W., 1990. Winning isn't everything: rats need only food deprivation and not food reward to efficiently traverse a radial arm maze. Learn. Motiv. 21, 153–163.
- Troyer, A.K., Murphy, K.J., 2007. Memory for intentions in amnestic mild cognitive impairment: time- and event-based prospective memory. J. Int. Neuropsychol. Soc. 13, 365–369.
- Tulving, E., 1972. Episodic and semantic memory. In: Tulving, E., Donaldson, W. (Eds.), Organization of Memory. Academic Press, New York.
- Tulving, E., 1993. What is episodic memory? Current Directions in Psychological Science 2, 67–70.
- Tulving, E., Markowitsch, H.J., 1998. Episodic and declarative memory: role of the hippocampus. Hippocampus 8, 198–204.
- Vargha-Khadem, F.G.D.G., 1997. Differential effects of early hippocampal pathology on episodic and semantic memory. Science 277, 376.
- Wilson, A.G., Crystal, J.D., 2012. Prospective memory in the rat. Anim. Cogn. 15, 349–358.
- Woods, S.P., Dawson, M.S., Weber, E., Gibson, S., Grant, I., Atkinson, J.H., 2009. Timing is everything: antiretroviral nonadherence is associated with impairment in time-based prospective memory. J. Int. Neuropsychol. Soc. 15, 42–52.
- Woods, S.P., Morgan, E.E., Marquie-Beck, J., Carey, C.L., Grant, I., Letendre, S.L., Group, T.H.N.R.C., 2006. Markers of macrophage activation and axonal injury

are associated with prospective memory in HIV-1 disease. Cogn. Behav. Neurol. 19, 217–221.

- Yin, H.H., Knowlton, B.J., 2004. Contributions of striatal subregions to place and response learning. Learn. Memory 11, 459–463.
- Zentall, T.R., 2005. Animals may not be stuck in time. Learn. Motiv. 36, 208-225.
- Zentall, T.R., 2006. Mental time travel in animals: a challenging question. Behav. Process. 72, 173–183.
- Zentall, T.R., 2010. Coding of stimuli by animals: retrospection, prospection, episodic memory and future planning. Learn. Motiv. 41, 225–240.
- Zentall, T.R., Clement, T., Bhatt, R., Allen, J., 2001a. Episodic-like memory in pigeons. Psychon. Bull. Rev. 8, 685–690.
- Zentall, T.R., Clement, T.S., Bhatt, R.S., Allen, J., 2001b. Episodic-like memory in pigeons. Psychon. Bull. Rev. 8, 685–690.
- Zentall, T.R., Singer, R.A., Stagner, J.P., 2008. Episodic-like memory: pigeons can report location pecked when unexpectedly asked. Behav. Process. 79, 93–98.
- Zhou, W., Crystal, J.D., 2009. Evidence for remembering when events occurred in a rodent model of episodic memory. Proc. Natl. Acad. Sci. U.S.A. 106, 9525–9952.
- Zhou, W., Crystal, J.D., 2011. Validation of a rodent model of episodic memory. Anim. Cogn. 14, 325–340.
- Zhou, W., Hohmann, A.G., Crystal, J.D., 2012. Rats answer an unexpected question after incidental encoding. Curr. Biol. 22, 1149–1153.

Contents lists available at SciVerse ScienceDirect





Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Monkeys show recognition without priming in a classification task

Benjamin M. Basile*, Robert R. Hampton

Emory University and Yerkes National Primate Research Center, Atlanta, GA, USA

ARTICLE INFO

Article history: Received 9 July 2012 Received in revised form 21 August 2012 Accepted 27 August 2012

Keywords: Implicit memory Nondeclarative memory Amnesia Repetition priming Categorization

ABSTRACT

Humans show visual perceptual priming by identifying degraded images faster and more accurately if they have seen the original images, while simultaneously failing to recognize the same images. Such priming is commonly thought, with little evidence, to be widely distributed phylogenetically. Following Brodbeck (1997), we trained rhesus monkeys (*Macaca mulatta*) to categorize photographs according to content (e.g., birds, fish, flowers, people). In probe trials, we tested whether monkeys were faster or more accurate at categorizing degraded versions of previously seen images (primed) than degraded versions of novel images (unprimed). Monkeys categorized reliably, but showed no benefit from having previously seen the images. This finding was robust across manipulations of image quality (color, grayscale, line drawings), type of image degradation (occlusion, blurring), levels of processing, and number of repetitions of the prime. By contrast, in probe matching-to-sample trials, monkeys recognized the primes, demonstrating that they remembered the primes and could discriminate them from other images in the same category under the conditions used to test for priming. Two experiments that replicated Brodbeck's (1997) procedures also produced no evidence of priming. This inability to find priming in monkeys under perceptual conditions sufficient for recognition presents a puzzle.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Seeing something makes it easier to see again. For example, if you saw a billboard partially obscured by a tree branch, you would likely identify the advertisement more readily if it were familiar than if it were novel. In the lab, this phenomenon is explored in studies of perceptual priming (sometimes called repetition priming; hereafter, priming; Tulving and Schacter, 1990). Subjects may be shown a set of novel images, the primes, and after a delay or manipulation, asked to name a number of degraded images, some that were seen previously, and others that are completely novel. Subjects typically show a memory effect in which they are more accurate or faster at naming the primed images than the unprimed images. Priming has been observed both with words, in which letters are missing (Tulving et al., 1982), and images, in which sections are missing or occluded (Mitchell, 2006; Snodgrass and Feenan, 1990).

A critical feature of theories of the organization of human memory is that priming is implicit. Subjects are better at identifying primed material even when they cannot explicitly remember that same material (Tulving and Schacter, 1990). Priming effects have

E-mail address: bbasile@emory.edu (B.M. Basile).

been found to last from days (Tulving et al., 1982), to weeks (Mitchell and Brown, 1988), to years (Mitchell, 2006), long after subjects have forgotten the original prime. Unlike our ability to recognize images, which improves over childhood, priming is relatively constant over development (Parkin and Streete, 1988). Priming is most strikingly dissociated from other types of memory in amnesic patients who show priming despite damage to their temporal lobes that prevents them from forming new explicit memories (Gabrieli et al., 1990; Hamann and Squire, 1997). Thus, priming is cognitively distinct from explicit memory, long-lasting, early developing, and depends on neural systems distinct from those serving explicit memory.

Because priming is automatic, effortless, unconscious, and neuroanatomically separate from sophisticated forms of memory such as episodic memory, researchers have theorized that it may be evolutionary old (Tulving, 1995). Additionally, a neural mechanism theorized to underlie priming, repetition suppression (Wiggs and Martin, 1998), has been observed in electrophysiological recordings in nonhuman primates (Mcmahon and Olson, 2007). This theory and evidence leads us to expect priming in monkeys.

It is therefore surprising that the only evidence of priming in nonhumans comes from pigeons (Brodbeck, 1997). Pigeons were trained to discriminate images of cats from images of cars and these images could be masked by placing small black boxes over a random 50% of the image. An irrelevant warning image preceded each trial. On study trials, the warning image was the S+ that would appear in the subsequent test trial. On test trials, the warning image was

^{*} Corresponding author at: Department of Psychology and Yerkes National Primate Research Center, 36 Eagle Row, Atlanta, GA 30322, USA. Tel.: +1 404 727 9619.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.08.005

normal, but the S+ had been seen as the warning image in a previous trial. Pigeons showed facilitated discrimination when the S+ image had been seen as the warning image on a previous trial, and this facilitation was evident even when the previous study trial occurred long enough ago that the pigeon would be unlikely to recognize the primed image. Brodbeck (1997) concluded that this facilitation was probably implicit and likely a result of priming. Oddly, the pigeons' accuracy was also facilitated on study trials, which the author suggests might represent short-lived conceptual priming.

There is one published claim of visual priming in monkeys based on evidence that monkeys were slightly faster at making symmetrical/asymmetrical judgments if they had recently seen the target stimulus (Mcmahon and Olson, 2007). Although this does show that prior exposure facilitated processing of the stimuli, no evidence was provided to indicate that the memory effect was implicit or otherwise dissociable from recognition performance. Given the short delays used (mean = 1.9 s), it is likely the monkeys would have recognized the primes and, thus, this study does not provide evidence for implicit priming in monkeys.

The phenomenon of perceptual priming is also similar to, but distinct from, the concept of "specific searching image" in studies of insect predation by birds (Tinbergen, 1960). After repeated exposure to cryptic prey, such as when foraging for camouflaged moths, birds may form a mental image that facilitates subsequent searches for that prey. However, unlike the facilitation caused by perceptual priming, the facilitation caused by forming a search image is short lived (Langley et al., 1996) and diminished by divided attention (Dukas and Kamil, 2001). This suggests that formation of a search image acts as a type of short-term attentional priming (Blough, 2001; Reid and Shettleworth, 1992; Shettleworth, 2010) rather than the long-term perceptual priming seen in humans.

We tested for priming in rhesus monkeys using procedures similar to those that have been effective in humans. As in priming studies in which humans were asked to name partially occluded images that they either had or had not seen previously (Mitchell, 2006; Snodgrass and Feenan, 1990), we required monkeys to classify partially occluded photographs that they either had or had not seen previously. Based on the similarity of this method to that used with humans, and the positive findings from pigeons (Brodbeck, 1997), we hypothesized that monkeys would show superior performance, either in accuracy or latency, for images they had previously seen compared to novel images. Had this been the case we would have needed to discriminate between true priming and contamination by memory processes supporting recognition. We would have conducted subsequent experiments to determine whether facilitation of classification occurred under conditions in which monkeys did not recognize the prime, which would indicate that priming is distinct from recognition and might suggest that the facilitation of classification occurred implicitly. But these further tests were not required because we found no evidence of facilitation of classification by previous exposure.

2. Experiment 1 - Initial results

2.1. Materials and methods

2.1.1. Subjects

We tested six adult male rhesus monkeys (*Macaca mulatta*; mean age at start of testing=3.2 years). They were pair-housed, on a 12-h light/dark cycle, received a full food ration daily, and had ad libitum access to water. All monkeys had prior experience with various cognitive tests using a touchscreen computer, including delayed matching-to-sample. Prior to this study, monkeys had learned to classify photographs as containing birds, fish, flowers, or people by touching one of four symbols. They all learned to a high-level of accuracy and transferred immediately to novel photographs without a significant decrement in accuracy (Paxton, Basile, Brown, and Hampton, submitted). During the course of this study, monkeys also participated in a variety of other cognitive tests, intermixed with the described experiments during a given day. During testing, monkeys were separated from their partners by a protected-contact divider that allowed the monkeys to see and groom each other, but not to access other individuals' testing rigs.

2.1.2. Apparatus

Subjects were tested six days a week in their home cages, using portable testing rigs controlled by Presentation testing software (Neurobehavioral Systems, Albany, CA). Each rig was equipped with a 15 in. color LCD touch-screen (3M, St. Paul, MN) running at a resolution of 1024×768 , generic stereo speakers, and two automatic food dispensers (Med Associates Inc., St. Albans, VT) that dispensed into food cups beneath the screen. Food rewards were nutritionally complete, banana flavored pellets (Bio-Serv, Frenchtown, NJ), supplemented by miniature chocolate candies on a random 10%of correct trials. One testing rig was attached to the front of each monkey's cage and remained there for 7 h, allowing the monkeys to work at their own pace during the day.

2.1.3. Stimuli

Stimuli were color photographs (Figs. 1 and 2a) gathered from the online photo repository Flickr (Yahoo!, Sunnyvale, CA) using the batch downloading tool FlickrDown (http://greggman.com). Duplicates were eliminated using DupDetector (Prismatic Software, Anaheim, CA) and visual inspection, were shuffled randomly and renamed using 1-4a Rename (http://www.1-4a.com), and were cropped to 300×400 pixels using Adobe Photoshop (Adobe, San Jose, CA). Each photograph depicted birds, fish, flowers, or people. For each photograph, the target category was the dominant subject of the image and each image could not depict multiple categories. Otherwise, images varied widely in perceptual features (e.g., fish could be in schools or alone, goldfish or sharks, in the ocean or on a plate). Because the priming effect should be easiest to measure when comparing previously-seen images to completely novel images, the size of the stimulus set was effectively unbounded, as novel images were added to the set each session.

2.1.4. Procedure

We ran one 100-trial session per day. Each session was preceded by presentation of primes. Monkeys saw each of 12 completelynovel images one at a time in the center of the screen and had to touch each one ten times (FR 10) to trigger presentation of the next image. We presented three images from each of the four categories in random order. One hundred classification trials immediately followed the 12 primes. To start a trial, monkeys touched a green box in the bottom center of the screen twice (FR 2). They then saw a single image in the center of the screen and touched it to bring up the four classification symbols. Touching the correct classification symbol resulted in food reward and positive audio feedback ("Woohoo!" or "Excellent!"). Touching one of the incorrect classification symbols resulted in a 5-s time out, during which the screen was black, and negative audio feedback ("D'oh!"). Trials were separated by an unfilled 3-s interval.

Of the 100 classification trials, the first four were always warmup trials using familiar unmasked images. Of the remaining 96 classification trials, each block of eight trials contained two images from each of the four categories, and half of the images were covered with a black checkerboard mask. The checkerboard mask was composed of black boxes (40 pixels wide × 30 pixels tall; Figs. 1 and 2a) that obscured 50% of the image. The four unmasked images were always familiar. Of the four masked images in each



Fig. 1. Diagram of a prime (left), and primed and unprimed classification trials (right). For primes at the beginning of the session, monkeys saw and touched (FR = 10) a novel, unmasked photograph. On classification trials, monkeys touched a green start box (FR = 2) to initiate the trial, saw and touched (FR = 2) a masked sample image, and finally classified the masked image by touching (FR = 2) one of the four colored symbols. The symbols used were: red triangle = birds, yellow star = fish, blue plus = people, green circle = flowers. On primed trials, the to-be-classified image had been seen once as one of the primes. On unprimed trials, the to-be-classified image was completely novel.

block of eight trials, two were familiar, one was a primed image from the initial 12 trials at the start of the session, and one was an unprimed image that was completely novel. Within each block of eight trials, the order of the images, which four images were masked, and the placement of the primed and unprimed images was random. Within the session, the primes were presented in the reverse of the order in which they had initially been seen, allowing us to assess the priming effect at a relatively wide range of delays within a single session. We ran twenty-four sessions, resulting in 288 primed trials and 288 unprimed trials for each monkey. For each session of this experiment and all subsequent experiments, the 12 primed and 12 unprimed images were novel. The remaining set of familiar images was continually refreshed by incorporating the 24 previously-novel images from the previous day's testing and eliminating the oldest 24 images.

2.1.5. Data analysis

For all experiments, accuracy was measured as the proportion of images correctly classified. All proportions were arcsine transformed prior to statistical analysis to better approximate normality (Aron and Aron, 1999). Latency was the median time from onset of the sample to selection of one of the symbols on correct trials only. Paired *t*-tests were used to compare performance on primed trials to that on unprimed trials, and one-sample *t*-tests were used to compare performance to chance levels. All statistical tests were two-tailed with α = .05.

2.2. Results and discussion

Monkeys were not more accurate at classifying primed images than unprimed images (Fig. 3, left; $t_{(5)} = 0.6$, p = .6). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 3, right; $t_{(5)} = 0.1$, p = .9). It is unlikely that the lack of memory effect was due to classification accuracy being at a level where facilitation would be undetectable. Human subjects showed the largest priming effect when stimulus identification accuracy was roughly halfway between ceiling and floor levels, about 60% (Snodgrass and Feenan, 1990). The absolute classification accuracy of our monkeys was roughly in this range as well, with room for priming to improve performance. Overall, this result surprised us, as the monkeys had seen the primed images unmasked at the start of the session, whereas they had never seen the unprimed images, masked or unmasked.

3. Experiment 2 - Immediate primes

In Experiment 1, we had hypothesized that we would observe a memory effect of having seen the primed images at the start of the session. However, we did not observe any memory effect at all. In Experiment 2, we attempted to increase memory for the primes by presenting them immediately before each primed trial, rather than at the beginning of the session. We hypothesized that reducing the delay between prime and classification of the masked primed image would increase memory for the prime and produce a memory effect.

3.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 1 (see Fig. 2a).

3.1.1. Procedure

Procedures were the same as in Experiment 1, with the exception that the initial 12 priming trials were interleaved with the classification trials, such that each prime was displayed immediately before the beginning of the trial on which the prime was to be classified. A 3-s intertrial-interval separated primes and primed trials. Thus the minimum time from seeing the intact prime to classifying the masked version of the same image was approximately 3 s. Unprimed trials were preceded by other classification trials, as in Experiment 1.

3.2. Results

Monkeys were not more accurate at classifying primed images than unprimed images (Fig. 4, left; $t_{(5)} = 1.2$, p = .3). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 4, right; $t_{(5)} = 1.2$, p = .3).

4. Experiment 3 - Black and white images

In Experiments 1 and 2, monkeys were no better at classifying recently-seen color photographs than at classifying completely novel color photographs. It is possible that this is because the monkeys were not classifying the photographs based on the shapes of the category exemplars depicted, but based on a simpler perceptual cue that correlated with category membership. In one classic study of classification by monkeys, for example, researchers attempted to train monkeys to classify color photographs based on the presence or absence of humans; however, the monkeys actually learned to



Fig. 2. Examples of stimuli used in Experiments 1–11. (A) Color photographs with black checkerboard masks, as used in Experiments 1, 2, and 10. (B) Black & white photographs with black checkerboard masks, as used in Experiments 3 and 4. (C) Black & white photographs with blur masks, as used in Experiments 5 and 6. (D) Line drawings with white checkerboard masks, as used in Experiments 7 and 8. (E) Line drawings with a white checkerboard mask in which the arrangement of un-occluded elements was either left intact or scrambled, as used in Experiment 9.



Fig. 3. Classification accuracy and response latency for primed and unprimed images in Experiment 1. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.



Fig. 4. Classification accuracy and response latency for primed and unprimed images in Experiment 2. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

classify based on the presence or absence of the color red, which happened to correlate with the presence or absence of humans (D'amato and Van Sant, 1988). In the current study, it seemed possible that our monkeys may also have learned a color discrimination; fish photos often had a blue background, humans photos usually contained flesh-tones, etc. In Experiment 3, we replaced the color photographs with black and white photographs. This tested whether monkeys were classifying images based on relatively simple color cues. Additionally, removing the color cues might increase the monkeys' attention to the shape of the category members depicted in each photograph. Human priming is thought to often be dependent on the shape of the primed image and our

⁽F) Color photographs of cats or cars with black masks composed of randomly-placed black squares, as used in Experiment 11a. (G) Color photographs of male or female rhesus monkeys with black masks composed of randomly-placed black squares, as used in Experiment 11b. For A-E, the same photograph is depicted to emphasize the various manipulations; however, primed and unprimed images were always novel for each experiment.



Fig. 5. Classification accuracy and response latency for primed and unprimed images in Experiment 3. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

ability to perceptually close occluded parts of that shape (Snodgrass and Feenan, 1990; Wiggs and Martin, 1998), so the priming effect in monkeys might be most evident when monkeys attend less to color and more to shape. We hypothesized that eliminating color cues would encourage attention to shape and produce a memory effect of having seen the unmasked image prior to classification.

4.1. Materials and methods

Subjects, apparatus, and procedures were as described in Experiment 2.

4.1.1. Stimuli

Stimuli were as described in Experiment 2 with the exception that all photographs were fully desaturated using Adobe Photoshop (Adobe Systems Incorporated, San Jose, CA) to eliminate all color (see Fig. 2b).

4.2. Results

Monkeys were able to correctly classify black and white photographs well above chance both when unmasked ($t_{(5)}$ = 26.0, p < .001) and when masked ($t_{(5)}$ = 12.2, p < .001). Monkeys showed a nonsignificant trend towards being more accurate at classifying primed images than unprimed images (Fig. 5, left; $t_{(5)}$ = 2.4, p = .06). They were not faster at correctly classifying primed images compared to unprimed images (Fig. 5, right; $t_{(5)}$ = 1.0, p = .4). Accurate discrimination of the black and white images indicates that the lack of a significant memory effect in the previous experiments was not due to the monkeys classifying the images based on color alone. However, if experience with black and white images resulted in increased attention to the shape of the category members, it did not produce significant priming.

5. Experiment 4 – Category repetition control

In Experiment 3, monkeys showed a nonsignificant trend towards being more accurate at classifying primed images compared to unprimed images. In Experiment 4, we evaluated whether this trend would replicate under more controlled conditions. In Experiments 2 and 3, the monkeys saw and touched an unmasked prime approximately 3 s before being required to categorize a



Fig. 6. Classification accuracy and response latency for primed and unprimed images in Experiment 4. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

masked version of the same image. In contrast, when monkeys classified masked unprimed images, the previous trial had been a normal classification trial. In Experiment 4, we introduced category control images immediately before the unprimed trials, in which the monkeys saw and touched an unmasked novel image from the same category as the upcoming unprimed image. These category control images controlled for any effect of increased attention or motivation caused by having the primed trial preceded by an image of the same category.

5.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 3 (see Fig. 2b).

5.1.1. Procedure

Procedures were the same as in Experiment 3 with the exception that the monkeys now had to touch a novel image prior to both primed and unprimed trials. Like the primes, these category control images were presented in the center of the screen, required ten touches, and were separated from the next trial by an ITI of 3 s. Category control images were always novel, unmasked images from the same category as the masked unprimed image that monkeys would classify on the subsequent trial, but unlike primes, where not the same image. We ran six sessions.

5.2. Results

Monkeys were not more accurate at classifying primed images than unprimed images (Fig. 6, left; $t_{(5)} = 0.9$, p = .4). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 6, right; $t_{(5)} = 0.4$, p = .7). This suggests that the non-significant trend seen in Experiment 3 was statistical noise rather than an actual effect.

6. Experiment 5 – Blur mask

In Experiments 1 through 4, images were made more difficult to classify by overlaying a black checkerboard pattern (Fig. 1). This introduced a regularity to the image that may have encouraged monkeys to classify them based on small patches of texture or small individual features rather than on the overall shape of the



Fig. 7. Classification accuracy and response latency for primed and unprimed images in Experiment 5. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

discriminanda. In Experiment 5, we masked images by blurring them rather than applying the checkerboard pattern. We hypothesized that the blur mask would encourage attention to shape and produce a memory effect of having seen the unmasked image.

6.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 5 (see Fig. 2c).

6.1.1. Procedure

Procedures were the same as in Experiment 5 with the exception that the checkerboard mask was changed to a blur mask. Blurred versions of each image were created using a Gaussian blur filter in Adobe Photoshop (Adobe Systems Incorporated, San Jose, CA). Blur strength was adjusted for each individual monkey prior to this experiment such that the blur mask brought overall classification accuracy down to approximately halfway between chance level and that monkey's maximum accuracy with unmasked images.

6.2. Results

Monkeys were significantly more accurate at classifying primed images than unprimed images (Fig. 7, left; $t_{(5)} = 4.4$, p < .01, d = 1.8). However, they were not faster at correctly classifying primed images compared to unprimed images (Fig. 7, right; $t_{(5)} = 0.5$, p = .6). Although the accuracy benefit seen with primed images was statistically significant, it should be interpreted with caution. It represents a relatively small absolute improvement (3.1%), and was the lone significant result of our study thus far. It is possible that this result is a Type I error. We investigated this possibility in Experiment 6 by attempting to reproduce the result.

7. Experiment 6 – Depth of processing

In Experiments 1 through 5, monkeys were required to touch the primes ten times to indicate that they had seen them. However, it is possible that monkeys were ignoring these images, perhaps touching the center of the screen without actually looking at the image. In Experiment 6, we ensured that monkeys attended to the primes and category control images by requiring them to classify those images in the same manner as with normal unmasked classification trials.



Fig. 8. Classification accuracy and response latency for primes, primed images, and unprimed images in Experiment 6. Left three bars: mean proportion (\pm SEM) of unmasked primes, masked primed images, and masked unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

We hypothesized that requiring the monkeys to attend to and process the primes would strengthen the small memory effect found in Experiment 5.

7.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 5 (see Fig. 2c).

7.1.1. Procedure

Procedures were the same as in Experiment 5 with the exception that we now required monkeys to classify image primes and category control images, as described for normal trials, instead of merely touching them.

7.2. Results

Monkeys accurately classified the unmasked primes at levels significantly above chance (Fig. 8, far left; $t_{(5)} = 40.2$, p < .001). However, monkeys were not more accurate at classifying primed images than unprimed images (Fig. 8, left; $t_{(5)} = 0.7$, p = .5). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 8, right; $t_{(5)} = 0.4$, p = .7). These results indicate that the monkeys were indeed attending to the primes, and that they had processed the primes enough to accurately classify them, but that this processing had no effect when classifying a masked version of the same image 3 s later. These results fail to replicate the memory effect observed in Experiment 5 under conditions in which the effect should have been strengthened, suggesting that the effect seen in Experiment 5 was a Type I error.

8. Experiment 7 - Line drawings

In Experiments 1 through 6, monkeys classified photographs. However, the majority of priming studies in humans use line drawings (Mitchell, 2006; Mitchell and Brown, 1988; Snodgrass and Feenan, 1990) or words (Hamann and Squire, 1997; Tulving et al., 1982), which are also composed of lines. To better match the conditions under which comparable priming experiments have been done in humans, we replaced the photographs with line drawings. Removing most color and texture cues might also increase the monkeys' attention to the shape of the category members depicted in



Fig. 9. Classification accuracy and response latency for primes, primed images, and unprimed images in Experiment 7. Left three bars: mean proportion $(\pm$ SEM) of unmasked primes, masked primed images, and masked unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

each image. We hypothesized that using line drawings would produce a memory effect of having seen the unmasked line drawing.

8.1. Materials and methods

Subjects and apparatus were as described in Experiment 6.

8.1.1. Stimuli

Stimuli were as described in Experiment 6 with the exception that all photographs were digitally transformed into line drawings (see Fig. 2d) using Adobe Photoshop (Adobe Systems Incorporated, San Jose, CA). The final result looked like detailed versions of the classic Snodgrass figures (Snodgrass and Corwin, 1988) that have been used successfully in several studies of human priming (Snodgrass and Feenan, 1990).

8.1.2. Procedure

Procedures were the same as in Experiment 6 with the exception that we used a white checkerboard mask similar to that used in Experiments 1 through 4. Because the stimuli consisted of black line drawings on white backgrounds, the white checkerboard mask was effectively invisible on certain areas of each picture, reducing the grid-like nature of the black checkerboard mask (see Fig. 2d).

8.2. Results

Monkeys accurately classified the unmasked primes at levels significantly above chance (Fig. 9, far left; $t_{(5)} = 34.1, p < .001$). Transfer of classification performance to line drawings indicates that the monkeys were not classifying images based on texture or grayscale cues alone, and were likely classifying based on global shape. However, monkeys were not more accurate at classifying primed images than unprimed images (Fig. 9, left; $t_{(5)} = 0.3, p = .8$). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 9, right; $t_{(5)} = 2.2, p = .1$). Using line drawings did not produce any memory effect.

9. Experiment 8 - Increased prime exposure

In Experiments 6 and 7, we required monkeys to classify the unmasked prime before classifying a masked version of the same



Fig. 10. Classification accuracy and response latency for primed and unprimed images in Experiment 8. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

image. Surprisingly, this additional processing did not improve subsequent classification. In Experiment 8, we took this reasoning to its logical extreme and required monkeys to classify each unmasked prime ten times in a single session before classifying the masked version of the image. In humans, the size of the priming effect increases with repetition of the prime (Wiggs et al., 1997). Consequently, we hypothesized that repeating each unmasked prime ten times would facilitate subsequent classification of the masked primed image.

9.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 7 (see Fig. 2d).

9.1.1. Procedure

Each session consisted of two phases. The first half of the session consisted of 120 trials in which monkeys classified 12 novel, unmasked primes 10 times each. Primes were pseudorandomized such that each block of 12 trials contained each prime image in a random order. The second half of the session consisted of 24 trials in which monkeys classified the 12 masked primes and 12 completely novel, masked unprimed images. The order of primed and unprimed trials was pseudorandomized such that each block of four trials consisted of two primed and two unprimed trials in random order. Ten sessions were run, resulting in 120 masked primed images and 120 masked unprimed images.

9.2. Results

Monkeys were not more accurate at classifying primed images than unprimed images (Fig. 10, left; $t_{(5)} = 0.7$, p = .5). Nor were they faster at correctly classifying primed images compared to unprimed images (Fig. 10, right; $t_{(5)} = 0.9$, p = .4). Surprisingly, classifying an unmasked Image 10 times did not improve subsequent classification of a masked version of that same image.

10. Experiment 9 – Scrambled classification

In Experiments 7 and 8, monkeys classified line drawings, suggesting that they were not classifying images based on color or texture alone. However, it is still possible that monkeys classified

1.0

0.8

images based on very small features, such as the sharp angle formed by a bird's beak or the rounded line formed by a person's eye. Priming in humans is thought to rely on the shape of the primed image and on perceptual closure, our tendency to mentally fill-in occluded areas of an image (Snodgrass and Feenan, 1990; Wiggs and Martin, 1998); therefore, monkeys may not show priming if they classify images based on hyper-local cues. Indeed, memory for a whole intact image may not help subsequent masked classification at all if that masked classification is based on the presence or absence of a small, specific feature. In Experiment 9, we tested whether monkeys were classifying masked images based on the overall shape of the target or on small, hyper-local features by scrambling the visible sections of the masked image (similar to Aust and Huber, 2001). If monkeys perceive the masked image as an occluded whole, then the position of the un-occluded sections should matter and scrambling them should reduce accuracy. In contrast, if monkeys focus on small, relatively simple features, then performance should be unaffected because the small features are still present and intact when the un-occluded sections are scrambled.

10.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiments 7 and 8 (see Fig. 2e).

10.1.1. Procedure

Each session consisted of two phases. In phase 1, monkeys classified 24 warm-up images, consisting of unmasked line drawings. In phase 2, monkeys classified unmasked line drawing images. masked line drawing images, and masked line drawing images in which the visible segments were scrambled. The visible segments were randomly scrambled with the restrictions that they must remain upright, could not remain in the same location, could not occupy an area that was masked, and must remain the same distance from the center of the image. In this way, small features and their distance from the center of the image were preserved, but the arrangement of those features relative to one another was disrupted. That is, local features were preserved but global structure was disrupted. Each block of eight trials contained four unmasked trials, two intact masked trials, and two scrambled masked trials. One session of 424 trials was run, providing 24 warm-up, 200 unmasked, 100 intact masked, and 100 scrambled masked trials. The same 100 images were used in the intact masked and scrambled masked trials, with half the images seen first intact and half seen first as scrambled. Thus, both intact masked images and scrambled masked images contained identical features that differed only in their arrangement (see Fig. 2e).

10.2. Results

Scrambling the masked images significantly decreased classification accuracy compared to intact masked images (Fig. 11; $t_{(5)} = 5.2$, p < .01, d = 2.1). This suggests that monkeys were not classifying images based on small, hyper-local features alone, and that this cannot explain the lack of a memory effect observed in previous experiments. This result is consistent with the idea that monkeys, like humans, perceptually close missing gaps in the image, and with previous findings that monkeys perceive lines as continuing through gaps (Feltner and Kiorpes, 2010) and behind occluding objects (Fujita, 2001).

11. Experiment 10 - Recognition of primes

In Experiments 2, 6, and 8, we attempted to ensure that monkeys remembered the prime image by moving it closer to the primed test, by having them classify it, and by having them classify it ten



Fig. 11. Classification accuracy for masked images that were intact or scrambled in Experiment 9. Bars depict mean proportion correct (±SEM). The dashed line represents the proportion correct that would be expected by chance.

times. In Experiment 10, we assessed whether monkeys actually did remember the primes by intermixing recognition trials with the normal classification, primed, and unprimed trials. We tested recognition for both unmasked and masked versions of the novel primes at the same memory interval as between primes and primed trials. Based on previous studies in which our monkeys remembered various types of stimuli over similar memory delays (Basile and Hampton, 2010, 2011), we hypothesized that monkeys would recognize the primes.

11.1. Materials and methods

Subjects, apparatus, and stimuli were as described in Experiment 2 (see Fig. 2a).

11.1.1. Procedure

Procedures were the same as in Experiment 6 with the exception that some primed trials were replaced by two-choice recognition tests. In a recognition test, monkeys first touched a green start box to initiate the trial, and then saw two images, one on the left center of the screen and one on the right center of the screen. One of the recognition images was the primed image from the preceding trial and the other image was a novel distractor from the same category. The correct screen location was pseudorandomized and balanced such that the primed image was presented on each side twice randomly within each block of four trials. On some recognition trials, the two options were both masked by the black checkerboard mask as in Experiments 1 through 4. Touching the primed image resulted in food reward and positive audio feedback ("Woo-hoo!" or "Excellent!"). Touching the novel image resulted in a 5-s time out, during which the screen was black, and negative audio feedback ("D'oh!"). Trials were separated by an unfilled 3-s interval. This general procedure was familiar to the monkeys, but they had never done matching with these stimuli or intermixed with classification trials.

Each session consisted of 172 trials: 4 unmasked warm-up images, 12 masked familiar images, 12 unmasked familiar images, 48 unmasked primes, 24 masked primed, 24 recognition, 24 unmasked category control images, and 24 masked unprimed images. We ran five sessions with the recognition images unmasked and five sessions with the recognition images masked. This resulted in a total of 240 primed, 240 unprimed, 120 unmasked recognition, and 120 masked recognition trials.



Fig. 12. Classification and recognition accuracy for primed and unprimed images in Experiment 10. Left two bars: mean proportion (\pm SEM) of masked primed and masked unprimed images correctly classified. Right two bars: mean proportion (\pm SEM) of unmasked primes and masked primes correctly recognized. The dashed line represents the proportion correct that would be expected by chance. Asterisks mark recognition accuracy that is significantly higher than chance.

11.2. Results

Monkeys accurately classified the unmasked primes at levels significantly above chance ($t_{(5)}$ =34.7, p <.001). However, monkeys were not more accurate at classifying primed images than unprimed images (Fig. 12, left; $t_{(5)}$ =0.4, p=.7). Nor were they faster at correctly classifying primed images compared to unprimed images ($t_{(5)}$ =0.1, p=.9). In contrast, monkeys did recognize the primed images significantly above chance both when unmasked and when masked (Fig. 12, right; unmasked: $t_{(5)}$ =4.7, p<.01, d=1.9; masked: $t_{(5)}$ =4.4, p<.01, d=1.8). These results indicate that the monkeys did remember the primed images at the time they were classifying the masked version of those images, but that remembering the image did not facilitate classification.

12. Experiment **11a** – Attempted replication of pigeon procedures

We designed Experiments 1 through 10 based on studies of priming in humans. However, the only positive evidence of priming in nonhumans comes from a study of pigeons (Brodbeck, 1997). In Experiment 11a we attempted to replicate the methods used with pigeons as closely as possible. We hypothesized that monkeys would show a priming effect similar to what has been observed with pigeons.

12.1. Materials and methods

Subjects and apparatus were the same as in Experiments 1 through 10.

12.1.1. Stimuli

Stimuli were 80 color photographs each of cats, cars, or stilllife images that contained neither cats nor cars. Half of each type of photo was used in training, and the other half was used for transfer and testing. All photos were gathered from the online photo repository Flickr (Yahoo!, Sunnyvale, CA) using the batch downloading tool FlickrDown (http://greggman.com). Duplicates were eliminated using DupDetector (Prismatic Software, Anaheim, CA) and visual inspection, were shuffled randomly and renamed using 1-4a Rename (http://www.1-4a.com), and were cropped to 400 × 400 pixels using Adobe Photoshop (Adobe, San Jose, CA). For each photograph, the target category was the dominant subject of the image and each image could not depict multiple categories. Otherwise, images varied widely in perceptual features (e.g., cats could be alone or in groups, outdoors or inside, running or sleeping).

12.1.2. Procedure

Procedures were closely based on those used successfully with pigeons (Brodbeck, 1997) except where noted.

Testing consisted of four phases. In phase 1, monkeys learned an S+/S- discrimination. On each trial, one cat and one car were presented on the center left and center right of the screen. The location of the S+ was balanced and pseudorandomized such that the S+ was presented twice on each side of the screen randomly within each block of four trials. If the monkey touched the S+ image, he received a food reward and a positive audio reinforcer ("woohoo!" or "excellent!"). If the monkey touched the S-, he received a negative audio reinforcer ("d'oh!"). All trials were separated by a 30-s ITI during which the screen was black. The S+ category was counterbalanced across monkeys such that cats were S+ for three monkeys and S- for the other three monkeys. Forty images were used from each category. One 200-trial session was run per day, for three days. This phase differs from that used with the pigeons in that stimuli were presented with a touchscreen rather than a slide projector, and that the 600 total trials were run in three sessions of 200 trials rather than 15 sessions of 40 trials.

In phase 2, we introduced a warning stimulus that monkeys had to touch to proceed to the discrimination. The warning stimuli were 40 still-life color photographs that contained neither cats nor cars. On each trial, one warning stimulus was presented in the center of the screen, monkeys touched it 20 times to proceed, the warning stimulus was replaced by a black screen for 200 ms, and then the discrimination trial proceeded as described in phase 1. One 200-trial session was run. This phase differed from that used with pigeons in that the 200 trials were presented in one session rather than five sessions of 40 trials.

In phase 3, we assessed whether monkeys had learned the category discrimination by transferring them to 40 novel cat, 40 novel car, and 40 novel warning images. All other procedures were the same as in phase 2. This phase differed from that used with pigeons in that the 200 trials were presented in one session rather than five sessions of 40 trials.

In phase 4, we assessed how much of the photographs would need to be occluded by the mask to produce an accuracy level near to that of the pigeons. Mask titration was necessary because we expected monkeys' classification accuracy to be much better than that of pigeons, and because priming would not be observable if accuracy was at ceiling level. Discrimination images were masked by randomly placing a number of black boxes (50×50 pixels) over each image (see Fig. 2f). Individual trials varied in how many blocks were placed, occluding 50%, 60%, 70%, or 80% of the image. One session of 200 trials was run, with 25 trials with each occlusion level. This phase was not present in the pigeon study; however, masks were constructed as described with the pigeons with the exception that they were made of pixels rather than electrical tape.

In phase 4, we tested for priming by sometimes replacing the warning stimulus with the S+ stimulus that would be seen on the subsequent trial. We ran three types of trials. On control trials, the warning stimulus and discrimination were presented as described in phase 3. On study trials, the warning stimulus was replaced with the S+ that would be seen on the subsequent trials. On test trials, the S+ was the warning stimulus from the previous trial. For example, on trial n, the warning stimulus might be cat#2 and the discrimination could be cat#1 versus car #1; then on trial n + 1, the warning stimulus might be a house and the discrimination could be cat#2. All discrimination images were masked at the level determined by phase 3; the warning stimuli were always unmasked. Each session contained 200 trials: 100 control, 50 study,



Fig. 13. Classification accuracy and response latency for primed and unprimed images in Experiment 11a. Left two bars: mean proportion $(\pm SEM)$ of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

and 50 test. Four sessions were run. This phase differed from that used with pigeons in that we ran four sessions of 200 trials rather than 16 sessions of 40 trials.

12.2. Results

Monkeys learned the cats/cars discrimination to over 80% accuracy by the end of the first training session and 98% accuracy by the final training session. They were 98% correct when transferring to novel stimuli. Monkeys required the highest mask level, which occluded 80% of the image and brought performance down to 81% during phase 3. In phase 4, monkeys were not more accurate at discriminating primed images on test trials than they were at discriminating unprimed images on control trials (Fig. 13, left; $t_{(5)} = 0.7$, p = .4). Nor were they faster at correctly discriminating the primed images than the unprimed images (Fig. 13, right; $t_{(5)} = 1.5$, p = .2). Under conditions similar to those in which pigeons showed a memory effect of having recently seen the to-be-classified stimuli (Brodbeck, 1997), monkeys did not.

13. Experiment 11b – Replication of 11a with species-relevant categories

In Experiment 11a, we attempted to reproduce the positive evidence of priming in pigeons (Brodbeck, 1997) using methods and stimuli that were as similar as possible. However, under these similar conditions, monkeys did not behave like pigeons. In Experiment 11b, we re-ran Experiment 11a with new stimuli. Instead of cats and cars, monkeys had to discriminate color photographs of adult male rhesus monkeys from photographs of adult female rhesus monkeys. These stimuli might produce a memory effect because overall discrimination performance should be further way from ceiling levels, as male and female monkeys are more perceptually similar than cats and cars, or because monkeys' perceptual systems may have evolved under pressures to process natural stimuli, such as fellow monkeys while obscured behind other objects. We hypothesized that monkeys would show a priming effect similar to what has been observed with pigeons.



Fig. 14. Classification accuracy and response latency for primed and unprimed images in Experiment 11b. Left two bars: mean proportion (\pm SEM) of masked primed and unprimed images correctly classified. The dashed line represents the proportion correct that would be expected by chance. Right two bars: group mean of the individual median response latencies to correctly classify primed and unprimed images.

13.1. Materials and methods

Subjects, apparatus, and procedures were the same as in Experiment 11.

13.1.1. Stimuli

Stimuli were 80 color photographs each of adult male rhesus monkeys, adult female rhesus monkeys, or still-life images that did not contain monkeys (see Fig. 2g).

13.2. Results

Monkeys learned the male/female discrimination to over 91% accuracy by the final training session and were 89% correct when transferring to novel stimuli, which included novel photographs of the same individuals and novel photographs of novel individuals. Monkeys required the highest or second-highest mask level, which occluded 80% or 70% of the image and brought performance down to an average of 72% during phase 3. In phase 4, monkeys were not more accurate at discriminating primed images on test trials than they were at discriminating unprimed images on control trials (Fig. 14, left; $t_{(5)} = 2.0$, p = .1). Nor were they faster at correctly discriminating the primed images than the unprimed images (Fig. 14, right; $t_{(5)} = 1.5$, p = .2). Using a paradigm similar to one that provided positive evidence of priming with pigeons (Brodbeck, 1997), and stimuli that should be ecologically relevant to monkeys, monkeys showed no benefit of having recently seen a to-be-classified image.

14. General discussion

Across 11 experiments, monkeys did not show priming, nor did they show any reliable benefit of having seen to-be-classified images. Our results clearly show that the monkeys attended to the primes, processed them enough to classify them accurately, and remembered them at the time they were classified for the second time. It is surprising that remembering images to a degree sufficient to support accurate recognition did not help monkeys classify them.

It is unlikely that our results were due to the monkeys classifying images based on relatively simple cues. Monkeys could not have learned specific responses to individual images, because they transferred classification performance to novel images without performance decrement (Paxton, Basile, Brown, and Hampton, submitted) and continued to accurately classify novel images throughout all experiments. Nor were monkeys classifying images based on relatively simple cues, such as color or texture, as the exemplars within each category varied widely in their perceptual features and monkeys continued to accurately classify images when in color, in black and white, line drawings, occluded by black squares, occluded by white squares, or blurred. Finally, monkeys were not classifying images based on small, hyper-local features alone, as shuffling the visible features significantly reduced accuracy. Together, these results indicate that the monkeys classified images based on global form, a process that would be expected to be facilitated by memory of the image.

Our null results are unlikely to be the product of low statistical power. First, it is unlikely that the size of our subject pool masked detection of priming. Priming is often seen in case studies of single human subjects (Gabrieli et al., 1990; Hamann and Squire, 1997), and the positive evidence from pigeons comes from a study using fewer subjects than the current experiment (Brodbeck, 1997). Second, it is unlikely that our measures were overly-noisy or that there were insufficient opportunities to observe priming. Each experiment used a large number of trials, which reduced measurement noise for each subject, and we conducted a large number of experiments, which provided many opportunities to observe priming. Third, our methods were sufficient to detect memory for the primes, as we were able to detect significant recognition in Experiment 10. Together, this suggests we would have detected an effect if one had been present.

We failed to replicate the findings of the single study of priming in nonhumans (Brodbeck, 1997). This apparent inconsistency has several potential explanations. First, priming may be a type of memory present in humans and pigeons, but not in monkeys; however, this seems phylogenetically unlikely.

Second, priming could be a type of memory unique to humans, and the positive results seen in pigeons may be the result of some other process that is present in pigeons but not monkeys, or some process that was recruited by Brodbeck's (1997) procedures but not ours. Pigeons have repeatedly failed to show evidence that they perceptually complete occluded figures (Fujita and Ushitani, 2005; Sekuler et al., 1996; Ushitani and Fujita, 2005; Ushitani et al., 2001), and manipulations that should affect classification based on global shape, such as scrambling the image, often do not impair classification in pigeons (Aust and Huber, 2001). Thus, it appears that unlike our monkeys, pigeons are more likely to process local rather than global features in classifications tasks. If monkeys and pigeons classify visual images using different processes, this may account for the difference in results. We are aware of only one other attempt to replicate Brodbeck (1997), and that work also failed to find priming in pigeons (S.L. Astley, personal communication, March, 30, 2010). More published data, especially from pigeons, are clearly needed to properly evaluate the possibility that evidence of priming in pigeons is due to some other process.

Third, priming could be a type of memory shared among humans, monkeys, and pigeons, but the current study was not designed appropriately to detect it. Although we based our methods on studies that have showed positive evidence of priming in humans and pigeons, and we explored a large number of reasonable manipulations, it is possible we failed to identify the critical conditions under which we would find priming, or indeed any memory effect on classification, in monkeys. This seems likely, based on the strong theoretical case that priming should be phylogenetically widespread (Tulving, 1995) and the strong mechanistic case that the likely neural underpinnings of priming have been observed in monkeys (Wiggs and Martin, 1998). Evidence of priming in monkeys may yet emerge from studies using different techniques, such as presenting primes extremely rapidly (Bar and Biederman, 1998) or measuring performance on primed images that are not degraded (Mitchell and Brown, 1988).

Whether or not monkey memory systems include a system for perceptual priming, our suite of findings indicates a dissociation. Experiment 10 showed that monkeys remembered the primes sufficiently to recognize them while simultaneously showing no facilitation of classification. This suggests that the visual processing necessary for recognition is cognitively encapsulated from that necessary for classification. It may be of interest to conduct further studies to better characterize the nature of this dissociation of visual memory and visual classification. For now, recognition without priming in monkeys presents a puzzle.

Acknowledgments

We thank Sarah E. Ward, Emily K. Brown, Dina P. Chou, Michael W. Williams, Jacey C. Jones, and Michael R. Ortiz for help with testing subjects and collecting stimuli. We thank Anthony A. Wright, Stephan B. Hamann, and Suzette L. Astley for insightful feedback on our results. This work was supported by NIH grant 1R01MH082819, NSF grant 0745573, the National Center for Research Resources P51RR000165, the Office of Research Infrastructure Programs/OD P510D011132, and by the Center for Behavioral Neuroscience under the STC Program of the NSF under Agreement IBN-9876754.

References

- Aron, A., Aron, E., 1999. Statistics for Psychology. Prentice Hall, Upper Saddle River, NJ.
- Aust, U., Huber, L., 2001. The role of item- and category-specific information in the discrimination of people versus nonpeople images by pigeons. Anim. Learn. Behav. 29 (2), 107–119.
- Bar, M., Biederman, I., 1998. Subliminal visual priming. Psychol. Sci. 9 (6), 464–469.
- Basile, B.M., Hampton, R.R., 2010. Rhesus monkeys (macaca mulatta) show robust primacy and recency in memory for lists from small, but not large, image sets. Behav. Process. 83 (2), 183–190.
- Basile, B.M., Hampton, R.R., 2011. Monkeys recall and reproduce simple shapes from memory. Curr. Biol. 21 (9), 774–778.
- Blough, P.M., 2001. Cognitive strategies and foraging in pigeons. In: Cook, R.G. (Ed.), Avian Visual Cognition. , Available from: http://www.pigeon.psy.tufts.edu/avc/pblough/
- Brodbeck, D.R., 1997. Picture fragment completion: priming in the pigeon. J. Exp. Psychol. Anim. Behav. Process. 23 (4), 461–468.
- D'amato, M., Van Sant, P., 1988. The person concept in monkeys (cebus apella). J. Exp. Psychol. Anim. Behav. Process. 14 (1), 43–55.
- Dukas, R., Kamil, A.C., 2001. Limited attention: the constraint underlying search image. Behav. Ecol. 12 (2), 192–199.
- Feltner, K.A., Kiorpes, L., 2010. Global visual processing in macaques studied using kanizsa illusory shapes. Vis. Neurosci. 27 (3–4), 131–138.
- Fujita, K., 2001. Perceptual completion in rhesus monkeys (macaca mulatta) and pigeons (columbia livia). Percept. Psychophys. 63 (1), 115–125.
- Fujita, K., Ushitani, T., 2005. Better living by not completing: a wonderful peculiarity of pigeon vision? Behav. Process. 69 (1), 59–66.
- Gabrieli, J.D.E., Milberg, W., Keane, M.M., Corkin, S., 1990. Intact priming of patterns despite impaired memory. Neuropsychologia 28 (5), 417–427.
- Hamann, S.B., Squire, L.R., 1997. Intact perceptual memory in the absence of conscious memory. Behav. Neurosci. 111, 850–854.
- Langley, C.M., Riley, D.A., Bond, A.B., Goel, N., 1996. Visual search for natural grains in pigeons (columba livia): search images and selective attention. J. Exp. Psychol. Anim. Behav. Process. 22 (2), 139–151.
- Mcmahon, D.B.T., Olson, C.R., 2007. Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. J. Neurophysiol. 97 (5), 3532–3543.
- Mitchell, D.B., 2006. Nonconscious priming after 17 years: invulnerable implicit memory? Psychol. Sci. (Wiley-Blackwell) 17 (11), 925–929.
- Mitchell, D.B., Brown, A.S., 1988. Persistent repetition priming in picture naming and its dissociation from recognition memory. J. Exp. Psychol. Learn. Mem. Cogn. 14 (2), 213–222.
- Parkin, A.J., Streete, S., 1988. Implicit and explicit memory in young-children and adults. Br. J. Psychol. 79, 361–369.
- Reid, P.J., Shettleworth, S.J., 1992. Detection of cryptic prey: search image or search rate? J. Exp. Psychol. Anim. Behav. Process. 18 (3), 273–286.
- Sekuler, A.B., Lee, J.a.J., Shettleworth, S.J., 1996. Pigeons do not complete partly occluded figures. Perception 25 (9), 1109–1120.
- Shettleworth, S.J., 2010. Cognition, Evolution, and Behavior, 2nd ed. Oxford University Press, Oxford; New York.
- Snodgrass, J.G., Corwin, J., 1988. Perceptual identification thresholds for 150 fragmented pictures from the snodgrass and vanderwart picture set. Percept. Mot. Skills 67 (1), 3–36.

- Snodgrass, J.G., Feenan, K., 1990. Priming effects in picture fragment completion – support for the perceptual closure hypothesis. J. Exp. Psychol. Gen. 119 (3), 276–296.
- Tinbergen, L., 1960. The natural control of insects in pinewoods. Arch. Néerlandaises Zool. 13 (3), 265–343.
- Tulving, E., 1995. Organization of Memory: Quo Vadis? MIT Press/Cambridge, MA, USA/London, England, UK.
- Tulving, E., Schacter, D.L., 1990. Priming and human-memory systems. Science 247 (4940), 301–306.
- Tulving, E., Schacter, D.L., Stark, H.A., 1982. Priming effects in word-fragment completion are independent of recognition memory. J. Exp. Psychol. Learn. Mem. Cogn. 8 (4), 336–342.
- Ushitani, T., Fujita, K., 2005. Pigeons do not perceptually complete partly occluded photos of food: an ecological approach to the pigeon problem. Behav. Process. 69 (1), 67–78.
- Ushitani, T., Fujita, K., Yamanaka, R., 2001. Do pigeons (columba livia) perceive object unity? Anim. Cogn. 4 (3), 153–161.
- Wiggs, C.L., Martin, A., 1998. Properties and mechanisms of perceptual priming. Curr. Opin. Neurobiol. 8 (2), 227–233.
- Wiggs, C.L., Martin, A., Sunderland, T., 1997. Monitoring frequency of occurrence without awareness: evidence from patients with alzheimer's disease. J. Clin. Exp. Neuropsychol. 19 (2), 235–244.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Angular amplitude matters: Exploring the functional relationship of geometric cue use by Clark's nutcrackers (*Nucifraga columbiana*)

Debbie M. Kelly*, James F. Reichert

Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada

ARTICLE INFO

ABSTRACT

Article history: Received 7 July 2012 Received in revised form 22 September 2012 Accepted 2 October 2012

Keywords: Clark's nutcracker Geometry Parallelogram Angular Functional relationship Four groups of Clark's nutcrackers (Nucifraga columbiana) learned to search for food hidden at one of two geometrically identical corners of a parallelogram-shaped enclosure. The corners of the enclosure projected either 40° and 140° angles or 60° and 120° angles. Tests using both rhomboid and rectangular enclosures examined whether birds had encoded angular amplitude and the length of walls, respectively. Cue conflict tests using a mirror-image of the training parallelogram reversed the relationship between wall length and corner angle, allowing for the examination of cue weighing. Furthermore, cue conflict tests which manipulated the angular amplitude allowed for the investigation of whether the encoding of angular information was similar among the training groups. Our results showed that nutcrackers encoded both angular amplitude and wall length. During cue conflict tests that maintained the training angular amplitudes, birds did not show a strong weighing of angular cues at a population-level but rather considerable individual differences were found. Finally, manipulating angular amplitude in the direction towards the unrewarded angle resulted in reduced weighing of angular cues whereas manipulating angular amplitude in the direction away from the unrewarded angle resulted in greater weighing of angular cues. In summary, our results support the importance of using multiple exemplars during training and testing to better understand the functional relationship between geometric cues during a spatial search task.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In order to understand the mechanisms underlying a particular behavior or cognitive process one must examine the phenomena using a range of measurements. Dr. Anthony Wright has championed this approach within the area of comparative cognition through his ongoing study of functional relationships. Functional relationships are measures of behavior across the continuum of a given variable as opposed to measurements obtained at a single point (Wright, 2010). Functional relationships are thought to more accurately reflect the way in which animals process information, where experiences do not occur in isolation but rather as a series of events (Wright, 2007). Using this approach, Wright et al. have revealed patterns of learning and memory that might otherwise have gone unnoticed. For example, when pigeons, monkeys and humans were tested for their memory of visual lists, these three species showed similar effects of primacy and recency;

* Corresponding author at: P401 Duff Roblin Building, 190 Dysart Road, Department of Psychology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada. Tel.: +1 204 474 9489.

E-mail address: Debbie_Kelly@umanitoba.ca (D.M. Kelly).

however, it was only through the manipulation of delay intervals between items that quantitative inter-species differences in list memory were found (Wright et al., 1985). Similarly, concept learning by pigeons and monkeys during a same/different task was only achieved through systematically increasing the number of training sets used (Wright, 2010). Thus, by allowing animals an opportunity to sufficiently experience abstract relations between stimuli, we find that several species are able to learn relational concepts. This approach has also been successfully applied to the spatial domain, where the use of multiple exemplars has been instrumental in showing relational encoding by animals. Indeed, when Kamil and Jones (1997) sought to explore relational learning in a species of food-storing bird (Clark's nutcracker), it was the principle of multiple training exemplars that allowed them to show that the birds could successfully learn the concept of a relative middle that exists between two landmarks.

Following the same paradigm used by Kamil and Jones (1997) with Clark's nutcrackers, Jones et al. (2002) employed a comparative approach to study geometric learning across bird species. The researchers used one species of non-food-storing columbid (pigeons, *Columba livia*) and two species of corvid, one a foodstoring species (Clark's nutcrackers, *Nucifraga columbiana*) and the other non-storing (jackdaws, *Corvus monedula*). The researchers

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.10.002

trained the birds to search for a food reward located at the center point between two discrete landmarks; the inter-landmark distance was varied across trials. Once the birds were accurately locating the reward during training trials, they were tested with novel inter-landmark distances (both interpolated and extrapolated distances). Pigeons and nutcrackers were able to accurately transfer to novel inter-landmark distances, whereas the difficult nature of the training task for the jackdaws makes the interpretation of their search behavior less clear. Although, both pigeons and nutcrackers showed qualitatively similar transfer, the nutcrackers showed more accurate performance compared to the pigeons. This study supports species differences in the use of geometric information during a spatial task, a finding likely driven by natural history.

Tommasi and Polli (2004) developed a different approach to study the encoding of geometric information by birds, through the modification of a reorientation paradigm devised by Cheng (1986). During Cheng's original task, disoriented rats were trained to locate a reward hidden at one corner of a fully enclosed rectangular environment. Although distinctive featural information was provided at each corner, and during the reference memory paradigm the location of the reward was stable with regards to the features, the rats required a considerable number of training trials to learn the task. Furthermore, when featural and geometric cues were placed in conflict during the affine transformation test, the rats showed strong reliance on the environmental geometry. This paradigm has become a staple procedure for examining the encoding of geometric information for reorientation by many species [e.g., pigeons (Kelly et al., 1998), chicks (Vallortigara et al., 1990), fish (Sovrano et al., 2002), rhesus monkeys (Gouteux et al., 2001), ants (Wystrach and Beugnon, 2009), and human children and adults (Hermer and Spelke, 1994, 1996); for a review see Cheng and Newcombe (2005)].

Tommasi and Polli (2004) modified the environment such that, instead of using a rectangular-shaped space, chicks were trained to locate a reward positioned in front of one pair of geometrically correct corners within a parallelogram-shaped arena. The shape of the environment provided two types of distinctive geometric cues: wall length (two short and two long walls) and angular amplitude (two 60° and two 120° corners). One group of chicks was trained to find a hidden food reward at the 60° corners which were subtended by a long wall to the left and a short wall to the right, whereas a second group was trained to find a hidden food reward at the 120° corners which were subtended by a long wall to the right and a short wall to the left. Once the chicks were accurately locating the hidden reward, the researchers administered transformation tests to examine which geometric properties the chicks had encoded. The shape of the environment was transformed to either a rhombus or a rectangle in order to investigate the encoding of angular amplitude and wall length information, respectively. Test results showed that the chicks had encoded wall length and angular amplitude, and were able to use either cue in isolation with similar facility. The chicks were also presented with a cue conflict situation during which the environment underwent a mirror transformation that placed angular amplitude and wall length cues in opposition. For example, during training the two corners that were derived from a long wall to the left and a short wall to the right always projected a 60° acute angle. However, during the mirror transformation test these corners now projected a 120° obtuse angle, and vice versa for the 120° corner. The purpose of the mirror transformation test was to place angular and wall information in direct conflict, thereby measuring the relative weighing of these two sources of geometric information by the chicks. Results showed that chicks that had been trained with acute angles preferred angular cues whereas chicks trained with obtuse angles preferred wall length cues. The results from this study show that, although both groups had encoded both wall length and angular amplitude, training history was critical for how the birds weighed the two types of geometry (although see Cheng and Gallistel, 2005 for an alternative interpretation of the data).

Recently, a comparative investigation using pigeons adopted the parallelogram-shaped enclosure paradigm to examine how adult pigeons encode the two types of geometric cues: angular amplitude and wall length (Lubyk and Spetch, 2012). The results from this study show that pigeons also encoded both cue types but, unlike chicks, they showed strong control by the angular amplitude, irrespective of training angle. This study raises the question whether the difference found between chicks and pigeons is due to developmental differences or species differences?

Recent studies with human participants have shown that how angles are encoded may be influenced by the amplitude of the angle used in training. For instance, Reichert and Kelly (2012) trained humans using an angle discrimination task to always select one angle associated with reinforcement when paired with a second angle (Experiment 1, or two additional angles during Experiment 2). During Experiment 1, two groups were trained to discriminate between a 50° and 75° angle. For group 50, choosing the 50° angle resulted in reinforcement whereas for group 75 selecting the 75° angle resulted in reinforcement. Both groups learned the task to high accuracy and were subsequently tested with their reinforced angle paired with a novel angle. The novel angle was selected from a range of angles smaller or larger than the training angles (i.e., 30° , 35° , 40° , 45° , 55° , 60° , 65° , 70° , 80° , 85° , 90° , and 95°). Overall, the researchers found that both training groups showed a preference to select the novel angle over the training angle when the amplitude of the novel angle was shifted away from the comparison training angle; thus, group 50 chose the smaller angles more frequently than the angle rewarded during training (50°) and group 75 chose the larger angles more frequently than the angle rewarded during training (75°). However, when examining the responses to these novel angles, the researchers found that group 75 showed relative encoding - they did not discriminate among the larger angles, whereas group 50 showed absolute encoding – the two angles nearest to the training angle (45° and 40°) were chosen more frequently than the smallest angles (35° and 30°). Thus, for humans, the amplitude of the angles used during training influenced whether participants encoded the angles using a relative or absolute metric. The aforementioned studies examining the use of angular amplitude and wall length by birds have not manipulated the angular amplitude, thus it is not known how birds encode the angular properties of the parallelogram enclosure.

Since all previous studies using the parallelogram-shaped enclosure paradigm have used non-food-storing birds as subjects, we wanted to examine how Clark's nutcrackers encode geometric information during this type of spatial search task. As noted earlier, Kamil et al. have shown that nutcrackers are more accurate when encoding a relative middle rule compared to pigeons and jackdaws. Thus, it is possible that food-storers encode and weigh geometric information differently than non-storers. We trained four groups of nutcrackers to search for a food reward consistently located at one corner of a parallelogram-shaped enclosure. Once the birds were accurately locating the reward during training, rectangle and rhombus shape transformation tests were administered to examine whether the birds had encoded both wall length and angular cues. Mirror transformation tests of the parallelogram enclosure were conducted to examine how the nutcrackers weighed these two types of geometric cues when the arrangement of the cues provided conflicting information regarding the rewarded location. Using a functional relationship approach, we examined cue reliance across a wide range of angular amplitudes while holding wall length stable.

2. Methods and materials

2.1. Subjects

Twenty wild-caught, adult Clark's nutcrackers (*N. columbiana*) participated in this study (ten male and ten female). All of the birds completed other unrelated experiments prior to this study. The birds were maintained at 85% of their ad libitum feeding weight with parrot pellets, sunflower seeds, pine nuts, mealworms and vitamin powder. Water and grit was freely available. The nutcrackers were kept in individual cages (73 cm height \times 48 cm width \times 48 cm depth) within a colony room. The colony room was kept on a 12:12-hr light-dark cycle, with light onset at 0700.

2.2. Apparatus

The parallelogram-shaped enclosure (140 cm length \times 70 cm width \times 60 cm height) was used for training. The size of this enclosure was constructed to be double the size of the parallelogram-shaped enclosure used by Tommasi and Polli (2004) to approximate differences in the relative subject-to-enclosure size. The enclosure was made of wood, with the inside walls painted gray such that all walls were of identical color. Each wall was attached to the next by piano hinges, allowing for 180° movement. A square porthole (25 cm) was centered in each wall to allow birds access into and out of the enclosure. The porthole was accessible through a guillotine-style door that was operated from outside of the enclosure. The enclosure was placed on a base (244 cm length \times 122 cm width) mounted on castors which allowed the entire apparatus to be moved within the experimental room.

Identical small tin containers were attached to the base of the enclosure with Velcro[®]; one container was located 10 cm from each corner. During all reinforced trials, two pine nuts were placed inside one of the containers whereas the other three were empty [Note: We only rewarded one of the two geometrically correct corners to be consistent with the methods of Tommasi and Polli (2004). This differs from the rewarding of both geometrically correct corners by Lubyk and Spetch (2012)]. The corner with the reinforced container was counterbalanced across birds within each group. The bottom of the enclosure was covered with approximately 5 cm of aspen wood chips. Two fluorescent lights, each measuring 60 cm in length, were attached to the long walls of the enclosure with metal clips to allow the lights to be removed from the walls. The light fixtures were interchanged within a session such that they could not provide a stable cue. A veil composed of ten layers of toile material (9 layers white, 1 layer black) was placed over top of the enclosure, allowing the experimenters to view the inside of the enclosure while preventing the birds from seeing out. This was possible as illumination only came from inside the enclosure (the larger experimental room was dark), so the toile veil appeared opaque from within the enclosure, but transparent from outside of the enclosure. A video camera was suspended over the top of the enclosure, which was attached to a Sony GVD digital recorder, which recorded all of the birds' choices for later scoring purposes. White noise was played through two speakers outside of the enclosure to mask any external noise; the position of these speakers was randomized across trials. A small holding container was used to transport the birds to and from the experimental enclosure.

A second enclosure was constructed to be identical to the parallelogram-shaped enclosure but was rhombus-shaped; all walls were of identical lengths and the same height as the parallelogram-shaped enclosure ($104 \text{ cm} \text{ length} \times 104 \text{ cm} \text{ width} \times 60 \text{ cm}$ height). All other aspects of this enclosure and the experimental set-up were identical. The size of this enclosure was also constructed to be double the size of the rhombus-shaped enclosure used by Tommasi and Polli (2004).

2.3. Procedures

2.3.1. Groups

The birds were divided into four groups of five birds each, counterbalancing for sex (two groups had 3 males and 2 females whereas the other two groups had 3 females and 2 males). Two groups were trained with the corners of the parallelogram set to angles of 40° or 140° (see Fig. 1a). The birds in Group 40 were trained to locate food at one of the two 40° corners whereas the birds in Group 140 were trained to locate food at one of the two 140° corners (the rewarded corner was counterbalanced across birds within both groups). The remaining two groups were trained with the corners of the parallelogram set to angles of 60° and 120° (see Fig. 1b.). The birds in Group 60 were trained to locate food at one of the two 60° corners whereas the birds in Group 120 were trained to located food at one of the two 120° corners (the rewarded corner was counterbalanced across birds within both groups). As this was a reference memory task, each bird's rewarded corner (herein referred to as the "positive corner") remained the same throughout the experiment.

2.3.2. General procedures

At the beginning of each day, the experimental enclosure was rotated 90° within the experimental room to reduce the chance that the birds were able to orient using external stimuli. Before each trial within a session, all tin containers were replaced, debris was removed from the apparatus and the wood shavings were smoothed. Before the first trial, an individual bird was placed into the holding container and transported from the colony room to the experimental room where it was placed in the rotation chamber. The bird was then slowly rotated for 12 revolutions for one minute. This procedure was used to disorient the bird so that inertial cues could not be used to locate the positive corner (Kelly et al., 1998). The rotation chamber was placed flush against one of the enclosure's four portholes (counterbalanced for order). The lights in the enclosure were illuminated and the guillotine door was raised, allowing the bird access to the inside. Once the bird completed the requirements of the trial (explained below) the lights in the enclosure were extinguished, and the light in the rotation chamber was illuminated, which encouraged the bird to exit the enclosure and enter the rotation chamber. The guillotine door was lowered, confining the bird to the rotation chamber where it remained until the enclosure was prepared for the next trial. At the end of the session, the bird was removed from the rotation chamber, placed in the holding container and returned to the colony room.

2.3.3. Training procedures

Training was divided into four phases. During the *first training phase*, only the container in the positive corner, for each bird, was present. Each daily session consisted of four trials. During each trial, the bird was given five minutes to approach and eat the pine nuts inside the container – this was considered a "choice". Once a bird was attaining reinforcement within one minute or less for each trial during one daily session, the bird began the second training phase.

During the *second training phase*, four identical uncovered containers were placed one in each corner of the enclosure. Only the container in the positive corner for each bird was reinforced with two pine nuts; the remaining three containers were empty. A choice was defined as the bird placing its beak into a container. From this phase onward, only the first two choices were recorded, although each bird was allowed to continue making choices until it chose the reinforced container. To advance to the next training phase, the bird had to make a choice within one minute or less, with the added requirement that each bird had to complete a minimum of two daily sessions.

During the *third training phase*, all of the containers were covered with a square piece of paper towel. Thus, the bird needed



Fig. 1. Schematic representations of the enclosures presented during training and testing. (a) Schematic representation of the parallelogram-shaped enclosure for groups 40 and 140. For illustrative purposes, the filled circles represent the location of the positive tin containers, located at the corners with 40° angular amplitudes, for Group 40, whereas the open circles represent the location of the positive tin containers, located at the corners with 40° angular amplitudes, for Group 40, whereas the open circles represent the location counterbalanced among birds in each group). (b) Schematic representation of the parallelogram-shaped enclosure for Groups 60 and 120. For illustrative purposes, the filled circles represent the location of the positive tin containers, located at the corners with 120° angular amplitudes, for Group 120. Only one of these positive containers was rewarded during training (the location of the positive tin containers, located at the corners with 120° angular amplitudes, for Group 120. Only one of these positive containers was rewarded during training (the location counterbalanced among birds in each group). (c) Schematic of the Rectangle test which maintained wall length but all corner angles were an identical 90°. (d) Schematic of the Rhombus test which maintained angular amplitude as during training but all the walls were an equal length. Thus, Group 40/140 experienced only the 40/140 Rhombus test and Group 60/120 experienced only the 60/120 Rhombus test. (e) Schematic of the Mirror parallelogram tests, the relative encoding of angular amplitude was examined.

to peck through the paper towel covering, or otherwise remove the covering, in order to determine whether or not the container held reinforcement. During the trials of the first session, the paper towel was placed loosely on top of the container. Only the positive corner was reinforced. For all subsequent sessions in this phase, a piece of paper towel was secured with an elastic band to the top of each container and the edges of the towel were trimmed. A choice was defined as the bird pecking through or otherwise removing the paper towel from a container. A choice was considered correct if it was made to either the positive corner or its geometrically equivalent corner. To advance to the fourth training phase, the bird had to make 26 or more first choices to the positive corner or the geometrically equivalent corner (depending on group assignment) across eight daily sessions (32 total trials).

During *Training phase four*, the final training phase, the number of training trials per day was increased to five. Two trials, but never the first, were not reinforced; during these non-reinforced trials all containers (including the one in the positive corner) were empty. Non-reinforced trials were included to familiarize the birds with trials that ended with non-reinforcement, as all testing conditions were conducted without reinforcement. For a bird to advance to testing, it had to make 32 or more first choices to the positive corner or the geometrically equivalent corner (depending on group assignment) across eight daily sessions (40 trials). When these criteria were met, the bird began the testing phase.

2.3.4. Testing

Each bird experienced five different testing conditions with each condition presented a total of five times. Each daily session consisted of three baseline trials, one control trial, and one test

trial. The order of the trials within a session was counterbalanced, with the exception that the first trial of each session was always a baseline trial. Baseline trials were identical to the reinforced trials during training. The enclosure was in the shape of a parallelogram (with angles set to the training angles) and only the positive corner was reinforced. Control trials were identical to baseline trials with the exception that they were non-reinforced. In addition, the procedures for preparing the control trials were made as similar as possible to that of the test trials; this was achieved by moving the enclosure into the testing configuration and then moving it back to the training configuration. During testing trials, the shape of the apparatus was transformed and no reinforcement was provided. Five types of tests were conducted in a randomized-block design and comprised of the Rhombus, Rectangle, Mirror parallelogram 40/140, Mirror parallelogram 60/120, and Mirror parallelogram 80/100 tests.

During the *Rectangle* test trials, the parallelogram enclosure was used but the shape was modified to form a rectangle (i.e., the amplitude of the corner angles were all set to 90°). During the Rectangle test, all angular information was equal, thus the only cue the birds could use to locate the two geometrically correct corners was the relative length of the four walls along with sense information (left-right relationship) learned from training (see Fig. 1c).

During the *Rhombus* test trials, the parallelogram enclosure was removed from the base and replaced with the rhombus enclosure. The lights, doors and veil from the parallelogram enclosure were attached to the rhombus-shaped enclosure. During the Rhombus test trials, the four walls of the enclosure were of equal length, thus the only cue the birds could use to locate the two geometrically correct corners was the angular amplitude from the enclosure's corners (see Fig. 1d).

To examine how the birds weighed the local cues (i.e., angular amplitude and wall length) the birds were presented with three types of cue conflict tests. These three tests used the parallelogram enclosure but it was reflected (mirrored) compared to training (see Fig. 1e). All birds, regardless of training group, experienced each of the Mirror parallelogram tests. Thus, during the *Mirror parallelogram 60/120* test, the only cue that changed for Groups 60 and 120 was the relationship between the angle and wall length (the absolute amplitude of the angles remained unchanged), whereas for Groups 40 and 140 the absolute value of the angular amplitude changed as well as the relationship between the relative angular amplitude and wall length. The opposite was true for the *Mirror parallelogram 40/140*. Finally, for the *Mirror parallelogram 80/100* the absolute amplitude changed as well as the sense relationship between angular and wall length for all groups.

In total, each bird completed 75 baseline trials 25 control trials and 25 test trials (five of each test type) over the course of testing.

2.4. Data recording and analysis

All non-reinforced control and test trials were videotaped. Only the first choice to a container was used in the following analyzes. All trials were re-scored by a second researcher naïve to the research hypothesis (inter-rater reliability was 99%), any differences were rescored by a third researcher naïve to the research hypothesis and this value was used for analyzes. One-way and mixed-model Analysis of Variance tests (ANOVAs) were used, significant *F* ratios were followed by *t*-tests (one sample and paired) or Tukey-Kramer Multiple Comparison tests to examine specific comparisons. All significance testing was conducted at α = 0.05.

3. Results

3.1. Training

A mixed-model ANOVA with between-subjects factor Group (40, 60, 120, and 140) and within-subjects factor Training Phase (3 and 4) was conducted to examine the percent of choices to the two geometrically correct corners during training. All groups showed similar accuracy which increased from training phase 3 to training phase 4 [$M \pm SE = 89.8\% \pm 4.2$, $83.6\% \pm 6.0$, $92.3\% \pm 4.4$ and $97.3\% \pm 9.7$, for Groups 40, 60, 120 and 140, respectively; F(3,16) = 3.10, p > 0.05; $M \pm SE = 83.0\% \pm 3.7$ and $98.6\% \pm 0.4$, for Phases 3 and 4, respectively; Phase: F(1,16) = 23.19, p < 0.001; Group × Phase: F(3,16) = 2.73, p > 0.07; see Fig. 2].

3.2. Control trials

A mixed-model ANOVA with between-subjects factor Group (40, 60, 120 and 140) and within-subjects factor Block (1–5) was conducted to examine the percent of choices to the two geometrically correct corners during control trials (5 trials per block). Choices were very accurate and no significant main effects or interactions were found [Group: F(3,16)=0.92, p>0.4; Block: F(4,64)=0.12, p>0.9; Group × Block: F(12,64)=1.04, p>0.4]. A One Sample *t*-test showed that choices to the two geometrically correct corners during the control trials for the birds within four groups were significantly greater than expected by chance [chance = 50%; $M \pm SE = 96.6\% \pm 1.017$; t(19)=45.83, p<0.001].

3.3. Rectangle and Rhombus tests

To examine whether the birds were able to use only angular amplitude or wall length information alone and whether the



Fig. 2. Average percent choice to the two geometrically correct corners for the four groups during training phases three and four. Error bars represent the S.E.M. Chance responding is 50%.

birds were able to use these two cue types with similar accuracy we conducted a mixed-model ANOVA with between factors Group (40, 60, 120 and 140) and Testing Condition (Rhombus and Rectangle). The birds in all groups showed similar accuracy during both tests and there was no significant interaction [Group: F(3,16)=0.68, p>0.5; Condition: F(1,16)=0.06, p>0.8; Group × Condition, F(3,16)=1.69, p>0.2; see Fig. 3]. A One Sample *t*-test (using the collapsed data) showed that choices to the two geometrically correct corners for Rhombus and Rectangle tests for the groups were significantly greater than expected by chance [chance = 50%; $M \pm SE = 86.5\% \pm 2.436$; t(19) = 14.98 p < 0.001]. Thus, the birds had encoded both the angular cues and the wall length during training and were able to use either cue with equal facility when presented alone.

3.4. Cue conflict tests – no amplitude manipulation

To examine how the birds would weigh angular amplitude and wall length cues when these cues provided conflicting information we conducted a one-way ANOVA with between factor Group (40, 60, 120 and 140) based on choices to the two geometrically correct corners according to angular amplitude. For this analysis, we used the data from the Mirror parallelogram tests for which there was no absolute change in angular amplitude for each of the groups; specifically, we used the data from Mirror parallelogram 60/120 for Groups 60 and 120 whereas we



Fig. 3. Percent of correct choices for each of the four groups during the Rhombus and Rectangle tests. Error bars represent the S.E.M. and **p < .01. Chance responding is indicated by the 50% bar.



Fig. 4. Percent of angular-based choices for each of the four groups during the Mirror parallelogram tests which maintained the same angles as during training (i.e., Mirror parallelogram 60/120 test for Groups 60 and 120, and Mirror parallelogram 40/140 test for Groups 40 and 140). Angular-based choices are compared to wall-based choices: **p < 0.01. Error bars represent the S.E.M. Equal cue use indicated by the 50% bar.

used the data from Mirror parallelogram 40/140 for Groups 40 and 140. This allowed us to examine only the effect of the cue conflict without any changes to the angular amplitude.

A significant main effect of Group [F(3,16) = 3.39, p < 0.05], which was driven by a significant difference between Groups 60 and 140, showed that group weighing of cue type differed during the cue conflict situation [angular choices: $M \pm SE = 72.0\% \pm 12.0$, $32.0\% \pm 20.591$, $68.0\% \pm 10.198$ and $92.0\% \pm 8.0$, for Groups 40, 60, 120 and 140, respectively. A Tukey–Kramer Multiple-Comparison test showed that the only significant difference was between Group 60 and Group 140; see Fig. 4].

However, paired *t*-tests showed that only Group 140 chose the corners correct according to angular amplitude significantly more than the corners correct according to wall length [ts(4) = 1.83, -0.87 and 1.77, all p > 0.1, for Groups 40, 60 and 120, respectively; whereas Group 140: t(4) = 5.25, p < 0.01]. Thus, at the group level, although Groups 40, 120 and 140 all showed a greater percentage of angular choices, this difference was only significant for Group 140.

Examining the individual choices of the birds within the groups, however, showed that several birds did weigh one type of cue more heavily than the other. Within Group 40, two birds relied upon angular amplitude; within Group 60, one bird relied upon angular amplitude, whereas three birds relied upon wall length information; for Group 120, two birds relied upon angular amplitude; finally, for Group 140, four birds relied upon angular amplitude. Thus, with the exception of Group 60, for the birds that did show a cue preference, it was the angular amplitude that was weighed more heavily than the wall length.

3.5. Cue conflict tests – amplitude manipulation

To examine how the birds in Groups 60 and 120 weighed their choices when one source of information remained constant whereas the other was changed, we held wall length constant and manipulated the angular amplitude by changing it by $\pm 20^{\circ}$. A Mixed-model ANOVA with between factor Group (60 and120) and within factor Degree of Change (-20° and $\pm 20^{\circ}$) was examined for choices to the two geometrically correct corners according to angular information. Although the groups did not differ overall [Group: *F*(1,8)=2.67, *p*>0.1], there was a main effect of Degree of



Fig. 5. Percent of angular-based choices for groups 60 and 120 during the Mirror parallelogram tests which changed the angular amplitude by -20° (cyan bars) or $+20^{\circ}$ (pink bars). The 0° responses during the Mirror parallelogram tests which did not change the angular amplitude are presented for comparative purposes (0°: white bars). Angular-based choices are compared to wall-based choices. Error bars represent the S.E.M. Equal cue use indicated by the 50% bar.

Change [F(1,8) = 14.40, p < 0.01)] as well as a significant interaction of Group × Degree of Change [F(1,8) = 67.60, p < 0.001; see Fig. 5].

Paired *t*-tests showed that, for Group 60, increasing the angular amplitude (+20°) caused the birds to significantly reduce their reliance on this cue [more reliance on wall-based choices, Ms = 16% and 84% for angles and walls, respectively; t(4) = -3.47, p < 0.05], whereas decreasing the angular amplitude (-20°) resulted in the birds dividing their choices equally between wall length and angular information [Ms = 56% and 44\%, walls and angles respectively; t(4) = -0.80, p > 0.4].

In comparison, for Group 120, increasing the angular amplitude (+20°) strengthened the birds' reliance on angular information, now making more angle based choices significantly greater compared to wall based choices [*Ms* = 84% and 16%, angles and walls, respectively; t(4) = -3.47, p < 0.05], whereas decreasing the angular amplitude (-20°) caused the birds to switch their strategy, now making more wall based choices compared to angle based choices [*Ms* = 8% and 92%, angles and walls, respectively; t(4) = -8.57, p < 0.01].

Groups 40 and 140 allowed us to examine cue weighing when the angular amplitude was modified by larger values. We examined how the birds in Group 40 responded when presented with angular information that changed by $+20^{\circ}$ and $+40^{\circ}$. Paired *t*-tests showed that changing the angular amplitude by the same absolute degree ($+20^{\circ}$) as with Groups 60 and 120, did not result in a cue preference [Ms = 48% and 52\%, angles and walls, respectively; t(4) = -0.20, p > 0.8], whereas changing the angular amplitude such that it was considerably larger ($+40^{\circ}$), resulted in the birds switching to a wall-based strategy [Ms = 72% and 28\%, walls and angles respectively; t(4) = -4.49, p < 0.05; see Fig. 6a].

For Group 140, paired *t*-tests showed that changing the angular amplitude to the same absolute degree (-20°) as with Groups 60 and 120, also did not alter cue preference; the birds in this group continued to show a strong reliance on angular cues [*Ms* = 92% and 8%, angles and walls, respectively; *t*(4)=8.57, *p*<0.01], whereas a larger change, making the angular amplitude considerably smaller (-40°) , caused the birds to shift from a strong angular preference to relying on angular amplitude and wall length cues equally [*Ms* = 44% and 56%, angles and walls, respectively; *t*(4)=-0.61, *p*>0.5; see Fig. 6b].



Fig. 6. (a) Percent of angular-based choices for Group 40 during the Mirror parallelogram tests which changed the angular amplitude by $+20^{\circ}$ (pink bars) or $+40^{\circ}$ (dark pink bars). The 0° responses during the Mirror parallelogram tests which did not change the angular amplitude are presented for comparative purposes (0°: white bars). Angular-based choices are compared to wall-based choices: *p < 0.05, **p < 0.01. Error bars represent the S.E.M. Equal cue use indicated by the 50% bar.

4. Discussion

Our results show that, similar to other bird species studied to date (pigeons: Kelly et al., 1998; chicks: Vallortigara et al., 1990), Clark's nutcrackers show strong control by geometry and are capable of using this cue alone to orient within a fully-enclosed environment lacking in featurally distinctive cues. Our results also show that the nutcrackers encoded both wall length information (sense and distance) and angular amplitude, and were able to use each cue in isolation with similar facility. Unlike other avian species, nutcrackers did not show a strong population-level preference for cue weighing. By examining the functional relationships, through the inclusion of two additional training groups (compared to previous studies) and manipulating the amplitude of the corner angles, we were able to extend our understanding of how nutcrackers weigh these two sources of geometric information.

4.1. Angles and wall length information

During our current study, when the enclosure was transformed from a parallelogram to a rectangle, all nutcrackers regardless of group directed the majority of their choices to the geometrically correct corners, thus showing that they could use the lengths of walls alone to reorient. Similarly, when the parallelogram was transformed into a rhombus (all walls of equal length) the birds chose corners that projected the correct angular amplitude, regardless of whether their training angle had been acute or obtuse. These findings show that the nutcrackers, similar to pigeons (Lubyk and Spetch, 2012) and chicks (Tommasi and Polli, 2004), had learned both angular and wall length information during training and could use either cue independently during testing.

4.2. Cue conflict and individual differences

Tommasi and Polli's (2004) study showed that, when wall length information and angular cues were placed in conflict, chicks' reliance on cue type was dependent upon initial training experience. Chicks trained with an acute angle (60°) weighed angular information more heavily, showing a preference to search at the corners containing the correct angular amplitude even though these corners were incorrect according to wall length cues. However, chicks trained with an obtuse angle (120°) instead showed a preference to search at the corner that contained the correct wall length information even though the angular amplitude was incorrect. Pigeons, regardless of training group, show strong control by angular information (Lubyk and Spetch, 2012).

Although pigeons and chicks show differential influence of initial training, the weighing of the cues by the individuals within a group was not variable – the weighing of cues was consistent at the level of the population (however, for an example of individual differences in visual discrimination by pigeons see Elmore et al., 2009). Unlike pigeons and chicks, the nutcrackers (with the exception of Group 140) did not show a strong population-level weighing of cue types. The results from the Rectangle and Rhombus tests provide clear evidence that all of the nutcrackers had encoded both wall length and angular amplitude. However, when the birds experienced a cue conflict situation, the nutcrackers showed strong individual differences. Although, many of the birds weighed angular amplitude more heavily, this preference was only significant at the group level for the birds trained with the most obtuse angle (140°).

The differences seen across studies, with pigeons and chicks showing strong population-level cue use whereas nutcrackers showing individual level cue use, are particularly interesting in light of previous research examining attentional side-biases during spatial tasks for these three species. Diekamp et al. developed an avian version of a spatial pseudoneglect "cancelation task" used for studying visuospatial side biases by humans (humans: Bowers and Heilman, 1980, and birds: Diekamp et al., 2005). During this task, a participant is seated at a table and presented with a sheet of paper containing several letters of the alphabet in a scattered formation. The participant is requested to cross-out (or cancel) as many instances of a particular letter (for example, the letter "A") within a set time limit. Research using this cancelation task has shown that healthy participants show a bias to over-select letters presented within the left side of space; this has been labeled the "left-sided bias" and is present at the level of the population (Jewell and McCourt, 2000).

Diekamp et al. modified the human cancelation task in order to study visuospatial biases by birds. Pigeons and chicks were trained initially to select grains that were presented in front of their body in a columnar arrangement. Once the birds were reliably selecting the grains, the experimenters presented the birds with a matrix of grains located centrally in front of the subject. Again, the birds simply needed to select the grains. The researchers found that, similar to human participants, both the pigeons and chicks showed a population-level bias to over-select grains located on their left side. This study was the first to show that birds have a left-sided attentional bias during the cancelation task (this leftsided bias in adult pigeons has been further replicated by Wilzeck and Kelly, 2012). However, adopting the same procedures as used by Diekamp et al., Kelly et al. found that Clark's nutcrackers did not show a left-sided population-level visuospatial attentional bias, but rather that many of the birds showed strong individual-level biases (Kelly, personal observation). These results may be important to our current study as, taken together, they suggest that how nutcrackers attend, and subsequently weigh, spatial information may be more variable compared to pigeons and chicks. The underlying mechanisms of this intriguing possibility are in need of empirical investigation.

4.3. Manipulating the amplitude – the importance of functional relationships

Our current study is the first to examine how systematic manipulations of angular amplitude during a re-orientation task influences cue weighing by birds. Generally speaking, our results show that manipulating the angular amplitude in the direction towards the unrewarded angle resulted in the birds weighing the angular cues less; conversely, manipulating the angular amplitude in the direction away from the unrewarded angle resulted in the birds showing a stronger weighing of angular cues.

For Group 60, when the angular amplitude was changed in the direction towards the unrewarded angle (making it more obtuse) the birds shifted from using angular amplitude and wall length information equally to predominately using wall length, whereas when the angular amplitude was shifted in the direction away from the unrewarded angle (making it more acute) the birds showed a slight increase in their reliance on angular amplitude. A similar pattern of results was found for Group 120, when the angular amplitude was changed in the direction towards the unrewarded angle (making it more acute) the birds shifted from using angular amplitude and wall length information equally to predominately using wall length, whereas when the angular amplitude was shifted in the direction away from the unrewarded angle (making it more obtuse) the birds showed a slight increase in their reliance on angular amplitude was shifted in the direction away from the unrewarded angle (making it more acute) the birds shifted from using angular amplitude and wall length information equally to predominately using wall length, whereas when the angular amplitude was shifted in the direction away from the unrewarded angle (making it more obtuse) the birds showed a slight increase in their reliance on angular amplitude.

It should also be noted that when manipulating the angular information towards the unrewarded angle for Groups 60 and 120 (making the angle more obtuse for group 60, now a 80° angle, or making the angle more acute for group 120, now a 100° angle) the four angles are also made more similar to that of the uninformative 90° angle from the Rectangle test (each angle only differs by ± 10 degrees from a 90° angle). Thus, for the Mirror parallelogram 80/100 tests the birds may have shifted their reliance away from using the angular cues, as they were less discriminable, to the more salient wall length cues.

Groups 40 and 140 showed parallel results to that of Groups 60 and 120 when large changes in angular amplitude were made in the direction away from the unrewarded angle. However, for these groups more change was necessary for the birds to shift cue weighing. Group 40 showed a systematic decrease, although non-significant, in the reliance on angular information as the amplitude shifted towards the unrewarded angle (more obtuse) whereas Group 140 shifted to include the use of wall information only with the largest change in the direction towards the unrewarded angle (more acute). Thus, our results show that manipulating the amplitude of the corner angles during testing had different effects on the groups, with the direction of the manipulation (towards or away from the unrewarded corner) affecting the birds' relative reliance on angular amplitude or wall length cues.

One possible interpretation of these results is that of adaptivecombination (Newcombe and Ratliff, 2007; Ratliff and Newcombe, 2008). Newcombe and colleagues suggest that cue use depends

on factors such as saliency and stability of the cue along with the subject's previous experience using the particular cue. At first glance our results seem to support this theory. Throughout testing, the birds experienced several instances during which the angular amplitude changed, thus possibly making amplitude an unstable cue compared to wall length which was only manipulated during the Rhombus test. It would be interesting to examine the functional relationship of wall length by replicating the current study but instead systematically manipulating wall length while holding angular amplitude constant (perhaps similar to Kelly and Spetch, 2001). This approach would not only provide more information regarding how cue stability affects the hierarchy of geometric cue use, but would also allow for a clearer understanding of whether wall length and angular information are indeed weighed equally (as suggested by the Rectangle and Rhombus tests). However, the explanation of cue stability may not account for the directionality of cue reliance; the direction of amplitude change, plus or minus 20°, would likely not affect cue stability differentially (unless the perception of these values is not equivalent). Yet, our results clearly show that the direction of change was an important factor for all groups.

The comparative approach to examining geometric cue use with this paradigm has certainly supported species differences in spatial cue use. Tommasi and Polli's study has shown that, for chicks, angular amplitude may be an important factor in cue saliency, with acute corners being more salient than obtuse corners; whereas pigeons do not show this difference but rather show a strong reliance on angular amplitude regardless of training experience. Clark's nutcrackers show yet another pattern of results. In general, angular amplitude is weighed heavily; however, nutcrackers show considerably more individual differences for the weighing of cues than either pigeons or chicks. However, as a group, the direction of angular change as well as initial training angle is important. Further research will need to examine what drives these species differences. It would be particularly interesting to examine whether other food-storing birds, or possibly only members of the family Corvidae, also show individual differences in the reliance on geometric cues.

Cheng and Gallistel (2005) provided an alternative explanation for the results reported by Tommasi and Polli (2004), suggesting that the chicks may have been using the principal axes of the enclosures to guide search behavior rather than local cues. Tommasi and Polli (2004) showed that chicks trained to locate food at either a 60° or a 120° corner of a fully-enclosed parallelogram-shaped enclosure had encoded both wall length information as well as angular amplitude. However, when the chicks were required to search in a Mirror parallelogram-shaped enclosure (causing a cue conflict situation) both groups of chicks searched at the acute corners. The authors provide an explanation of this seemingly contradictory result by suggesting that the acute angles provided a more distinct cue, compared to the obtuse angles, for the chicks. Cheng and Gallistel, however, argue that a more parsimonious explanation for these data is that during the transformation test the chicks followed along the major principal axis of the enclosure and searched at the nearest corner; this alternative explanation explains the group differences. However, recently Kelly and colleagues have shown that neither pigeons nor chicks show reliance on principal axes, but instead use medial axes and/or local geometric cues (Kelly et al., 2011). Our results from the cue conflict tests provide further, although indirect, support for the use of local geometric cues rather than principal axes. During the cue conflict situations, we would not expect individual search pattern differences by the birds had they been relying only on principal axes. Additionally, we would not expect that by systematically manipulating the angular amplitude in the direction away from the unrewarded angle would result in all groups showing a stronger weighing of angular cues, rather if the nutcrackers were using principal axes we would expect all groups to search at the acute angles – as illustrated by Cheng and Gallistel.

Our current study adopted the approach of functional relationships to examine the weighing of geometric cues by Clark's nutcrackers through the use of several training groups as well as through the selection of several instances of angular amplitude for testing. By presenting more than one exemplar at training and testing we were able to build a better understanding as to how manipulating angular information changes the weighing of geometric cues. The approach of examining functional relationships is one that Dr. Anthony Wright has supported, with great success, over the years. The results from our current study show that investigating the functional relationships, within and across cue types, will be an invaluable approach for understanding species differences that exist in the hierarchal weighing of geometric cues as well as other spatial information.

Acknowledgements

This research was supported by a Discovery grant to D.M.K. from the Natural Sciences and Engineering Research Council of Canada. We thank Dason Harker for assistance with data collection, Fei Peng for assistance with data scoring and Cordt Euler for assistance with constructing the apparatuses.

References

- Bowers, D., Heilman, K.M., 1980. Pseudoneglect: effects of hemispace on a tactile line bisection task. Neuropsychologia 18, 491–498.
- Cheng, K., 1986. A purely geometric module in the rat's spatial representation. Cognition 23, 149–178.
- Cheng, K., Gallistel, C.R., 2005. Shape parameters explain data from spatial transformation: comment on Pearce et al. (2004) and Tommasi & Polli (2004). J. Exp. Psychol.: Anim. Behav. Process 31, 254–259.
- Cheng, K., Newcombe, N.S., 2005. Is there a geometric module for spatial reorientation? Squaring theory and evidence. Psychon. B Rev. 12, 1–23.
- Diekamp, B., Regolin, L., Güntürkün, O., Vallortigara, G., 2005. A left-sided visuospatial bias in birds. Curr. Biol. 15, R372–R373.
- Elmore, L.C., Wright, A.A., Rivera, J.J., Katz, J.S., 2009. Individual differences: either relational learning or item-specific learning in a same/different task. Learn. Behav. 37, 204–213.

- Gouteux, S., Thinus-Blanc, C., Vauclair, J., 2001. Rhesus monkeys use geometric and nongeometric information during a reorientation task. J. Exp. Psychol. Gen. 130, 505–519.
- Hermer, L., Spelke, E., 1994. A geometric process for spatial reorientation in young children. Nature 370, 57–59.
- Hermer, L., Spelke, E., 1996. Modularity and development: the case of spatial reorientation. Cognition 61, 195–232.
- Jewell, G., McCourt, M.E., 2000. Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. Neuropsychologia 38, 93–110.
- Jones, J.E., Antoniadis, E., Shettleworth, S.J., Kamil, A.C., 2002. A comparative study of geometric rule learning by nutcrackers: (*Nucifraga columbiana*), pigeons (*Columbia livia*), and jackdaws (*Corvus monedula*). J. Comp. Psychol. 116, 350–356.
- Kamil, A.C., Jones, J.E., 1997. The seed-storing corvid Clark's nutcracker learns geometric relationships among landmarks. Nature 390, 276–279.
- Kelly, D.M., Chiandetti, C., Vallortigara, G., 2011. Re-orienting in space: do animals use global or local geometry strategies? Biol. Lett. 7, 372–375.
- Kelly, D.M., Spetch, M.L., 2001. Pigeons encode relative geometry. J. Exp. Psychol.: Anim. Behav. Process 27, 417–422.
- Kelly, D.M., Spetch, M.L., Heth, C.D., 1998. Pigeons' encoding of geometric and featural properties of a spatial environment. J. Comp. Psychol. 112, 259–269.
- Lubyk, D.M., Spetch, M.L., 2012. Finding the best angle: pigeons (*Columbia livia*) weight angular information more heavily than relative wall length in an open-field geometry task. Anim. Cogn. 15, 305–312.
- Newcombe, N.S., Ratliff, K.R., 2007. Explaining the development of spatial reorientation: modularity-plus-language versus the emergence of adaptive combination. In: Plumert, J., Spencer, J. (Eds.), The Emerging Spatial Mind. Oxford University Press, New York, pp. 53–76.
- Ratliff, K.R., Newcombe, N.S., 2008. Reorienting when cues conflict: evidence for an adaptive-combination view. Psychol. Sci. 19, 1301–1307.
- Reichert, J.F., Kelly, D.M., 2012. Discrimination of geometric angles by adult humans. Anim. Cogn. 15, 271–284.
- Sovrano, V.A., Bisazza, A., Vallortigara, G., 2002. Modularity and spatial reorientation in a simple mind: encoding of geometric and nongeometric properties of a spatial environment by fish. Cognition 85, B51–B59.
- Tommasi, L., Polli, C., 2004. Representation of two geometric features of the environment in the domestic chick (*Gallus gallus*). Anim. Cogn. 7, 53–59.
- Vallortigara, G., Zanforlin, M., Pasti, G., 1990. Geometric modules in animals' spatial representations: A test with chicks (*Gallus gallus domesticus*). J. Comp. Psychol. 104, 248–254.
- Wilzeck, C., Kelly, D.M., 2012. Avian visual pseudoneglect: the effect of age and sex on visuospatial side biases. In: Csermely, D., Regolin, L. (Eds.), Behavioral Lateralization in Vertebrates. Springer-Verlag, Berlin, pp. 55–70.
- Wright, A.A., 2007. An experimental analysis of memory processing. J. Exp. Anal. Behav. 88, 405–433.
- Wright, A.A., 2010. Functional relationships for determining similarities and differences in comparative cognition. Behav. Process. 85, 246–251.
- Wright, A.A., Santiago, H.C., Sands, S.F., Kendrick, D.F., Cook, R.G., 1985. Memory processing of serial lists by pigeons, monkeys, and people. Science 229, 287–289.
- Wystrach, A., Beugnon, G., 2009. Ants learn geometry and features. Curr. Biol. 19, 61–66.
Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Beacons and surface features differentially influence human reliance on global and local geometric cues when reorienting in a virtual environment

Kent D. Bodily^{a,*}, Zachary A. Kilday^a, Caroline K. Eastman^b, Katherine A. Gaskin^b, April A. Graves^b, Jonathan E. Roberts^b, Bradley R. Sturz^{a,*}

^a Georgia Southern University, GA, USA ^b Armstrong Atlantic State University, GA, USA

ARTICLE INFO

Article history: Received 1 July 2012 Received in revised form 17 September 2012 Accepted 28 September 2012

Keywords: Global geometry Local geometry Reorientation Beacons Surface features

ABSTRACT

In the reorientation literature, non-geometric cues include discrete objects (e.g., beacons) and surfacebased features (e.g., colors, textures, and odors). To date, these types of non-geometric cues have been considered functionally similar, and it remains unknown whether beacons and surface features differentially influence the extent to which organisms reorient via global and local geometric cues. In the present experiment, we trained human participants to approach a location in a trapezoid-shaped enclosure uniquely specified by global and local geometric cues. We explored the role of beacons on the use of geometric cues by training participants in the presence or absence of uniquely-colored beacons. We explored the role of surface features on the use of geometric cues by recoloring two adjacent walls at the correct location and/or adding a line on the floor which corresponded to the major principal axis of the enclosure. All groups were then tested in novel-shaped enclosures in the absence of unique beacons and surface features to assess the relative use of global and local geometric cues. Results suggested that beacons facilitated the use of global geometric cues, whereas surface features either facilitated or hindered the use of geometric cues, depending on the feature.

Published by Elsevier B.V.

Multiple cues are available to mobile organisms attempting to determine orientation with respect to their environment. For example, geomagnetic fields, celestial bodies, and vestibular cues have been identified as contributing to orientation ability (for a review, see Healy, 1998). In the reorientation literature, these spatial cues have generally been categorized as either geometric (e.g., angles, distances) or non-geometric cues (e.g., objects, colors; for a review, see Gallistel, 1990). One pervasive method for investigating these types of reorientation cues involves training subjects to locate a corner in a rectangular enclosure marked by distinct beacons. Following training, researchers often manipulate the shape of the environment and/or the location of the beacons to determine the extent to which reorientation relied on geometric and non-geometric cues (for a review, see Cheng and Newcombe, 2005, 2006; Tommasi et al., 2012).

Initially, geometric cues appeared to be considered a single functional class of spatial cues. Global geometric cues, such as the major principal axis of space (which passes through the centroid and approximate length and width of the entire space,

E-mail addresses: kdbodily@georgiasouthern.edu (K.D. Bodily), bradleysturz@georgiasouthern.edu (B.R. Sturz).

respectively), and local geometric cues, such as wall lengths and corner angles, were not considered as independent cues for reorientation (e.g., see Cheng, 1986; Miller and Shettleworth, 2007). This unitary classification of geometric cues appeared to occur because of the difficulty in isolating the use of global versus local geometric cues for reorientation. Recently, Bodily et al. (2011) developed a reorientation task that allowed for the discrimination between the use of global and local geometric cues. Specifically, two groups of human participants were rewarded for searching in an isosceles trapezoid-shaped enclosure at locations that maintained the reliability of local geometric cues (wall lengths and corner angles) across groups but manipulated the reliability of global geometric cues (i.e., the major principal axis) between groups. Specifically, one group was rewarded for searching only at the right-hand side of the major principal axis whereas another was rewarded for searching at both the left- and right-hand sides of the major principal axis. When tested in a rectangle, a parallelogram, and the parallelogram's mirror equivalent, group differences emerged with respect to the geometric cues utilized for reorientation. The group trained with unreliable global but reliable local geometric cues exclusively reoriented using local geometric cues. In contrast, the group trained with reliable global and local geometric cues reoriented using both global and local geometric cues. Our present understanding, then, is that humans (and other animals) are able to reorient via global geometric cues (e.g., Bodily

^{*} Corresponding authors at: Department of Psychology, Georgia Southern University, P.O. Box 8041, Statesboro, GA 30460, USA.

et al., 2011; Sturz et al., 2011) and local geometric cues independently (e.g., Bodily et al., 2011; Lubyk et al., 2012), and that global and local geometric cues are best categorized as separate functional classes.

Similarly, we argue that conceptualizing non-geometric cues as a single functional class is potentially problematic for the investigation of spatial reorientation. Non-geometric cues, often broadly referred to as featural cues (e.g., Cheng and Newcombe, 2005), include both surface-based features (e.g., colors, textures, and odors), which are directly tied to a surface of the enclosure (e.g., a blue wall), and discrete objects (e.g., beacons, landmarks), which are placed inside or outside of the enclosure. Perhaps the tendency to group surface-based features and discrete objects into a single class was due to their functional similarity-both types of cues may be utilized to reorient. However, categorizing all non-geometric cues as a single functional class may fail to make potentially important distinctions between these different types of cues. For example, systematic investigations of the differential influence of intramaze and extramaze cues in the place learning literature revealed functional differences between cue types which helped account for divergent findings in the literature (e.g., Brown and Bing, 1997; Olton and Collison, 1979; see also Babb and Crystal, 2003). Similarly, investigating how surface-based features and discrete objects differentially affect reorientation may reveal important differences which may shed light on divergent findings and advance theoretical accounts of spatial reorientation.

As James (1890) pointed out over a century ago, psychological processes work "under conditions; and the quest of the conditions becomes the psychologist's most interesting task" (p. 3; for a similar charge, see Cheng, 2008). Although conditions such as environment size (e.g., Learmonth et al., 2002; Ratliff and Newcombe, 2008; Sturz et al., 2012) and environment shape (e.g., Sturz et al., 2011; Sturz and Bodily, 2011) have been shown to influence reliance on global and local geometric cues, questions regarding the effects of beacons and surface-based features on the use of global and local geometric cues remains unclear (e.g., Cheng and Gallistel, 2005; Cole et al., 2011; McGregor et al., 2006; Pearce et al., 2006; Pecchia and Vallortigara, 2012). The motivation of this study was to further illuminate the conditions under which global and local geometric cues are used. In particular, we attempted to uncover potential similarities and differences regarding the influence of surface-based features and beacons on reorientation by global and local geometric cues

In the present experiment, we utilized the methodology established by Bodily et al. (2011) to explore the extent to which beacons and surface-based features may influence the use of global and local geometric cues for reorientation. The purpose of this study was two-fold. First, we investigated the extent to which the presence of surface-based features or beacons separately influenced the use of global and local geometric cues. Second, we investigated the extent to which combinations of surface-based features and beacons influenced the use of global and local geometric cues.

We trained eight groups of human participants, in a virtual environment, to respond to a location in an isosceles trapezoidshaped enclosure (see Fig. 1). The use of the trapezoid-shaped enclosure is of critical importance. First, like a kite, an isosceles trapezoid is rotationally asymmetric. That is, opposite corners of an isosceles trapezoid are made up of different local geometric cues. Second, like a rectangle, the axis of symmetry bisects opposite sides (whereas the axis of symmetry of a kite bisects opposite angles). That is, the principal axis of space bisects opposite sides of an isosceles trapezoid, allowing a meaningful comparison between performance in an isosceles-trapezoid-shaped enclosure and performance in a rectangular enclosure with regard to dependence on global geometric cues. Finally, these characteristics allowed us to

Table 1

Predicted allocation of responses to geometrically-correct locations relative to chance performance (0.5) by cue type.

	Testing enclosure			
Cue type used to orient	Rectangle	Parallelogram 1	Parallelogram 2	
Global cues Local cues Global & local cues	Above Equal Above	Above Above Above	Above Below Equal	

train participants to approach a corner that was uniquely specified by a combination of global (i.e., right-hand side of the major principal axis of space) and local (i.e., short wall left and right, and obtuse corner angle) geometric cues, such that responding could come to depend on global, local, or both types of geometric cues.

To explore the effect of beacons on the use of geometric cues, we trained participants either in the presence or absence of uniquelycolored beacons at each response location. To explore the effect of surface-based features on the use of geometric cues, we trained participants in one of four different Surface-Feature conditions: None, Walls, Floor, or Both. In the None condition, we did not add unique surface-based features to any surface of the enclosure. In the Walls condition, we recolored the two adjacent walls at the correct corner to be a darker shade of gray than the opposite walls. In the Floor condition, we added a black line on the floor which corresponded with the major principal axis of the enclosure (see Krider et al., 2001). In the Both condition, we recolored the black line on the floor.

Following training, we tested participants in the absence of unique beacons and surface-based features in the trapezoid and three additional enclosures. The trapezoid test enclosure allowed us to assess whether participants depended primarily on global or local geometric cues when the non-geometric cues were removed. The other test enclosures allowed us to assess dependence on global geometry in isolation (i.e., the rectangle enclosure), in alignment with the trained local geometric cues (i.e., the parallelogram 1 enclosure) or in conflict with the trained local geometric cues (i.e., the parallelogram 2 enclosure). As shown in Table 1, the predicted allocation of responses to the top-right and bottom-left corners (henceforth referred to as geometrically-correct locations) in each test enclosure depends on which cues are used to reorient. For example, as the rectangle enclosure does not contain the trained local geometric cues (e.g., 120° corner angle), responding at the geometrically-correct locations is predicted only if global cues are used. Alternatively, as the parallelogram 2 enclosure contains trained local geometric cues at incongruent corners relative to the training trapezoid, responding at the geometrically correct corners is predicted if only global geometric cues are used, while responding to the opposite corners (top-left and bottom-right) is predicted if only local geometric cues are used. Overall, this design allowed us to assess the extent to which conditions involving beacons and surface-based features influenced the relative use of global and local geometric cues to reorient.

1. Method

1.1. Participants

One hundred twelve undergraduate students (48 males and 64 females) served as participants. Participants received extra classcredit or participated as part of a course requirement.

1.2. Apparatus

An interactive, dynamic three-dimensional virtual environment was constructed and rendered using Valve Hammer Editor and



Fig. 1. Sample images from the first-person perspective (*top*) of the virtual-environment training enclosures appear above the schematics of training trials for each group. Below these are shown the schematics of the testing enclosures experienced by all participants. For illustrative purposes, the gray quad-arrows mark the position where participants entered the virtual enclosures for all training and testing trials. Large unfilled dotted circles indicate training location for all groups. Small, filled circles represent response locations where colored spherical beacons were visible for participants in the Beacons-Present conditions but not for participants in the Beacons-Absent conditions. Dotted lines represent the major principal axis of space for each testing enclosure (calculated using FreeMat v4.1 software).

run on the Half-Life Team Fortress Classic platform. A personal computer, 21-in. flat-screen liquid crystal display (LCD) monitor, gamepad joystick, and speakers served as the interface with the virtual environment. The monitor (1680 × 1050 pixels) provided a first-person perspective of the virtual environment (see top panels, Fig. 1). Participants used the joystick on the gamepad to navigate within the environment. Speakers emitted auditory feedback. Experimental events were controlled and recorded using Half-Life Dedicated Server on an identical personal computer.

1.3. Stimuli

Dimensions are long wall(s) \times short walls \times height and measured in virtual units (vu; 1 vu = \sim 2.54 cm). Texture brightness is reported in luma, a standard measure of the brightness

of a digital image which may range from 0 (black) to 255 (white). Four virtual enclosures were created (see Fig. 1): Trapezoid $(550 \times 275 \times 260 \text{ vu})$, Rectangle $(550 \times 275 \times 260 \text{ vu})$, Parallelogram 1 $(550 \times 275 \times 260 \text{ vu})$, and Parallelogram 2 $(550 \times 275 \times 260 \text{ vu})$. Corner angles for the trapezoid-shaped enclosures were 60° for both acute angles and 120° for both obtuse angles. Corner angles in the parallelograms were also 60° for both acute angles. Corner angles for the rectangle were 90° . Please note that all short walls shown in Fig. 1 were identical in length. All wall surfaces were textured with light-gray "concrete" [brightness (in luma): M = 187.7, SD = 7.19; Adobe Photoshop 12.1], the floor surface was textured with gray tile, and the ceiling was black. Depending on the condition, during training the adjacent walls at the goal location were textured with a dark-gray "concrete" [brightness (in luma): M = 113.9, SD = 7.08], and/or a solid black line (width = 8 vu) was applied to the floor. For half of the participants, enclosures contained four spherical, semi-transparent beacons ($48 \times 48 \times 48 vu$). During training, a red, blue, yellow or green beacon appeared consistently in the same corner of the enclosure, hovering above the floor (40 vufrom center of beacon to floor). During testing, all beacons were semi-transparent white. Please note that the small size of the beacons, relative to the walls of the enclosure, and their semitransparency prevented the beacons from obstructing the corners of the enclosure.

1.4. Procedure

Participants were informed to navigate to the location that transported them to the next virtual room and to move via the joystick on the gamepad: \uparrow (forward), \downarrow (backward), \leftarrow (rotated view left), and \rightarrow (rotated view right). Participants selected a location by walking into it. Selection of the rewarded location resulted in auditory feedback (bell sound) and a 7-s inter-trial interval (ITI) in which the monitor went black and participants progressed to the next trial. Selection of a non-rewarded location resulted in different auditory feedback (buzz sound) and required participants to continue searching.

1.4.1. Training

Training consisted of 12 trials. Participants were randomly assigned to a combination of Beacon-Status (Present or Absent) and Surface-Feature (None, Walls, Floor, or Both) conditions, thereby creating eight groups (see Fig. 1). Gender and number of participants were balanced across groups. In the Beacons-Present condition, the response locations (i.e., each corner of the enclosure) were marked with uniquely colored spherical beacons. In the Beacons-Absent condition, response locations were unmarked. In the None Surface-Feature condition, no unique surface features were added to the walls or floor. For the Walls Surface-Feature condition, the two adjacent short walls at the rewarded corner location were textured with a darker gray "concrete" than the other two walls. For the Floor Surface-Feature condition, a solid black line was added on the floor which ran the length of the enclosure and corresponded to the enclosure's major principal axis. For the Both Surface-Feature condition, both the recolored walls and the solid black line were present.

For all groups, the rewarded location was always located in the top-right corner (see Fig. 1) such that searching at the egocentric right-hand side of the major principal axis and at a location specified by short wall left, short wall right, and obtuse angle was rewarded. Participants began each trial in the center of the Trapezoid enclosure (marked with a quad arrow in Fig. 1) with a heading that was randomly selected from 0° to 270° in increments of 90° .

1.4.2. Testing

Testing consisted of 60 trials composed of 12 five-trial blocks. Each trial block was composed of four Training trials and one Test trial. Training trials presented during Testing were identical to those experienced during Training for each group (see above). The order of the Training and Test trials was randomized within each block. For each Test trial, one of four enclosures was presented: Trapezoid, Rectangle, Parallelogram 1, or Parallelogram 2. Each enclosure was presented once without replacement until all four had been presented. Each enclosure was presented three times (total of 12 test trials). Participants made one response during Test trials which resulted in no auditory feedback followed by the 7-s ITI and progression to the next trial. Participants began each Testing trial in the center of the enclosures (marked with quad arrows in Fig. 1) with a heading that was randomly selected from 0° to 270° in increments of 90°. For all participants, the surface features and unique beacons were absent during test trials. That is, all of the walls were the same color and the floor was free of the black line for all participants. For participants in the Beacons-Present condition, the uniquely colored beacons were all replaced with white beacons. For participants in the Beacons-Absent condition, the response locations remained unmarked, as they were in training (see Fig. 1).

2. Results

2.1. Training

Acquisition performance was measured by coding the first location that a participant visited (i.e., the first response) in each trial as being either correct (i.e., the rewarded location) or incorrect (i.e., one of the other 3 locations), and then by calculating the proportion correct per two-trial block. A three-way mixed analysis of variance (ANOVA) on acquisition performance with Block (1-6), Beacon Status (Present, Absent), and Surface Feature (None, Walls, Floor, Both) as factors revealed main effects of Block, F(5, 520) = 46.30, *p* < .001, Beacon Status, *F*(1, 104) = 49.26, *p* < .001, and Surface Feature, F(3, 104) = 2.75, p < .05. The Beacon Status × Surface Feature interaction was also significant, F(3, 104) = 3.92, p = .01. None of the other interactions were significant, Fs < 1.29, ps > .26. Overall, acquisition performance increased across blocks, as confirmed by a significant linear component of the trend analysis on the Block factor, *F*(1, 104) = 131.22, *p* < .001, suggesting that all groups improved across training blocks.

Fig. 2 plots the Surface-Feature × Beacon Status interaction. As shown, the interaction was due to equivalent performance across Surface-Feature conditions in the presence of beacons, and differences in performance across Surface-Feature conditions in the absence of beacons. These results were confirmed by one-way ANOVAs on overall acquisition performance with Surface Feature (None, Walls, Floor, Both) as a factor, conducted separately for each Beacon Status condition. For Beacons-Present, there was no effect of Surface Feature, F(3, 52) = 0.17, p = .92. However, for Beacons-Absent, there was a main effect of Surface Feature, F(3,52)=4.68, p<.01. In the absence of beacons, the None condition was significantly different from the Walls condition (Tukey's post hoc test, p < .01), but no other comparisons were significantly different (ps > .06). Lastly, we compared Beacons-Present and Beacons-Absent groups within each Surface-Feature condition. The Beacons-Present performed significantly better than the Beacons-Absent groups in the None, Floor, and Both Surface-Feature conditions, independent samples *t*-tests, *ts*(26) > 2.3, *ps* < .03. However, the Beacons-Present and Beacons-Absent groups did not differ in the Walls Surface-Feature condition, independent-samples ttest, t(26) = 1.42, p = .17. Overall, these results suggest that the presence of beacons facilitated acquisition, and that, in the absence of beacons, the shaded walls facilitated acquisition.

Fig. 3 (left panels) shows the acquisition performance across two-trial training blocks for Beacons-Present and Beacons-Absent conditions, plotted separately for each Surface-Feature condition. All groups learned to approach the correct corner by the third block of training. These results were confirmed by comparing the acquisition performance of each group to chance (0.25) across training blocks 3–6, one-sample *t*-tests, *ts* (13) > 2.33, *ps* < .037. Additionally, all groups reached asymptotic performance by the fourth block of training. Acquisition performance in training block 6 did not differ from training blocks 4 or 5 for any group, paired-sample *t*-tests, *ts* (13) < 1.88, *ps* > 0.08.

To summarize, the results from training suggest that the presence of unique beacons at each response corner facilitated acquisition independently from surface features. In the absence

Overall First-Response Accuracy in Training



Fig. 2. Overall mean proportion of correct first responses throughout training across Surface-Feature conditions by Beacons-Present (filled) and Beacons-Absent (unfilled) groups. Dashed lines represent chance (0.25). Error bars represent standard errors of the means.

of unique beacons, surface feature which uniquely specified the correct location (i.e., the Walls condition) also facilitated acquisition. Finally, all of the groups reached and maintained a consistent, above-chance, level of accuracy throughout the last three blocks of training. This result is particularly important, as above-chance acquisition performance is necessary in order to interpret performance on test trials.

2.2. Testing

Test trials assessed the extent to which responding depended upon global and/or local geometric cues. To this end, test enclosures were devoid of the added surface features and, for the Beacons-Present groups, all beacons were white. First, we assessed whether groups successfully reoriented in the absence of beacons and/or surface features by analyzing the distribution of responses to all four corners (i.e. response locations) of the trapezoid testing enclosure. Additionally, each corner of the trapezoid testing enclosure had a unique combination of geometric cues in common with the trained corner (i.e., top-left: no common cues, top-right: global and local cues, bottom-left: global cues, bottom-right: local cues). By analyzing the proportion of responses allocated to each corner, we assessed whether participants depended on global and/or local geometric cues to reorient.

Next, to further assess the extent to which responding depended upon global and local geometric cues, we analyzed the mean proportion of responses allocated to the geometrically-correct locations (i.e., top-right and bottom-left corners) across all testing enclosures. Lastly, to assess the combined effects of beacons and surface features on the use of global and local geometric cues, we conducted planned comparisons analyzing the mean proportion of responses allocated to the geometrically-correct locations of each testing enclosure across Beacons-Present and Beacons-Absent groups within each Surface-Feature condition. Table 1 summarizes how the use of each type of geometric cue (i.e., global, local, global and local) is predicted to affect the allocation of responses to the geometrically-correct corners in each testing enclosure.

2.2.1. Trapezoid testing enclosure

Fig. 3 (right panels) shows the mean proportion of responses to each corner (i.e., response location) of the trapezoid testing enclosure for Beacons-Present and Beacons-Absent conditions, plotted separately for each Surface-Feature condition. For each group, the number of responses to each response location (i.e., TL, TR, BL, BR) across the three trapezoid testing trials was analyzed via χ^2 goodness-of-fit tests against an expected uniform distribution (.25) to assess the extent to which participants reoriented in the absence of unique beacons and/or surface features. The Beacons-Present and Beacons-Absent groups in the None, Walls, and Floor Surface-Feature conditions were oriented in the trapezoid test enclosure, χ^2 s (3, N=42)>13.24, ps < 0.01. Additionally, one-sample *t*-tests (α = .05) were conducted to determine the corners at which the mean proportion of responses differed from chance (see Fig. 3, right panels). Although all of these groups were significantly oriented in the trapezoid test enclosure, differences in the distributions of responses suggest that the presence of beacons and/or surface features during training influenced the extent to which global and local geometric cues were used in testing.

In the Both Surface-Features condition (Fig. 3, bottom-right panel), the distribution of responses of the Beacons-Present and Beacons-Absent groups did not significantly differ from an expected uniform distribution, $\chi^2 s(3, N=42) < 6.57$, ps > 0.08. These results may suggest that the presence of both surface features, regardless of the presence or absence of beacons, may have hindered the use of global and local geometric cues. However, both groups equally distributed their responses among the top-right (global and local cues), bottom-left (global cues), and bottom-right (local cues) locations and allocated significantly fewer responses to the top-left location (no congruent cues) than expected by chance, one-sample *t*-tests, *ts*(13)>2.33, *ps*<.04. These results may suggest that participants were oriented, and that responding depended upon both global and local geometric cues equally. Their results on the other testing enclosures revealed which of these two interpretations is best.

2.2.2. All testing enclosures

A three-way mixed ANOVA on mean proportion of responses to the geometrically-correct locations (i.e., top-right and bottomleft) with Beacon Status (Present, Absent), Surface Feature (None, Walls, Floor, Both), and Enclosure Type (Trapezoid, Rectangle, Parallelogram 1, Parallelogram 2) as factors, revealed main effects of Surface Feature, F(3, 104) = 2.86, p < .05, and Enclosure Type, F(3,312) = 26.45, p < .001. Additionally, the Beacon Status × Surface Feature, F(3, 104) = 4.07, p < .01, Beacon Status × Enclosure Type, F(3,312) = 4.57, p < .01, and the Surface Feature × Enclosure Type, F(9,312) = 2.24, p < .05 interactions were significant. Neither the effect



Fig. 3. Training acquisition and trapezoid-test response distributions for Beacons-Present (filled) and Beacons-Absent (unfilled) groups plotted separately by Surface-Feature conditions. *Left panels* plot mean proportion of correct first responses across training in two-trial blocks. *Right panels* plot mean proportion of responses across corners (response locations) of the trapezoid-test enclosure in the absence of surface features and unique beacons. Dotted borders represent measures that do not differ from chance. Dashed lines represent chance performance (0.25). Error bars represent standard errors of the means.

of Beacon Status nor the interaction of Beacon Status × Surface Feature × Enclosure Type was significant, Fs < 1, ps > .41. Separate follow-up analyses were conducted to determine the source of these interactions.

Fig. 4 (top panel) shows the Beacon Status \times Surface Feature interaction. As shown, the interaction was due to the

Beacons-Absent group allocating more responses to the geometrically-correct locations than the Beacons-Present group in the Floor Surface-Feature condition, independent samples *t*-test, t(26) = 2.48, p = .02. There was no effect of the Beacon Status in the other three Surface-Feature conditions, ts(26) < 1.83, ps > .078. These results suggest that the presence of beacons may have



Testing Performance across Surface Features by Beacon Status







Fig. 4. *Top Panel.* Mean proportion of responses to the geometrically-correct locations (i.e., top-right and bottom-left) during testing for the Beacons-Present (filled) and Beacons-Absent (unfilled) groups plotted across Surface-Feature conditions (i.e., the Beacon Status × Surface Feature interaction). *Middle Panel.* Mean proportion of responses to the geometrically-correct locations during testing for the Beacons-Present and Beacons-Absent groups plotted across testing enclosures (i.e., Beacon Status × Enclosure Type interaction). *Bottom Panel.* Mean proportion of responses to the geometrically-correct locations during testing for each Surface-Feature condition plotted across testing enclosures (i.e., Surface Feature × Enclosure Type interaction). Dashed lines represent chance (0.5). Error bars represent standard errors of the means.

weakly facilitated the use of global geometric cues in the None, Walls, and Both, but not in the Floor, Surface-Feature conditions.

Fig. 4 (middle panel) shows the Beacon Status \times Enclosure Type interaction. As shown, the interaction was due to the Beacons-Absent groups allocating significantly more responses to the

geometrically-correct locations in the Parallelogram 1 enclosure and significantly fewer responses to the geometrically-correct locations in the Parallelogram 2 enclosure than the Beacons-Present groups, independent samples *t*-tests, ts(110) > 2.05, ps < .05. However, there was no effect of Beacon Status in the Trapezoid or Rectangle enclosures, independent samples *t*-tests, ts(110) < 0.86, *ps* > .05. These results suggest that the absence of beacons appears to have facilitated the use of local geometric cues when they were aligned (i.e., the Parallelogram 1 enclosure) or conflicted (i.e., the Parallelogram 2 enclosure) with global geometric cues. Interestingly, participants were still able to reorient via global geometric cues (i.e. above chance in the Rectangle enclosure).

Fig. 4 (bottom panel) shows the Surface Feature × Enclosure Type interaction. As shown, the interaction was due to significant differences between Surface-Feature conditions in the Rectangle and Parallelogram 2 enclosures. These results were confirmed by separate one-way ANOVAs with Surface Feature (None, Walls, Floor, Both) as the factor for each testing Enclosure Type. There was a main effect of Surface Feature in the Rectangle, F(3, 108) = 4.03, p < .01, and in the Parallelogram 2 enclosure, F(3, 108) = 2.78, p < .05, but not in the Trapezoid or Parallelogram 1 enlcosures, Fs(3, 108) < 1.76, ps > .15. In the Rectangle enclosure, participants in the Floor Surface-Feature condition responded more to the geometrically-correct locations than participants in the None and Both Surface-Feature conditions, Tukey post hoc tests, ps < .05. In the Parallelogram 2 enclosure, participants in the None and Walls Surface-Feature conditions appear to have allocated fewer responses to the geometrically-correct locations than participants in the Floor and Both Surface-Feature conditions, although no statistically significant differences were revealed, Tukey post hoc tests, *ps* > .10. These results suggest that the surface features influenced the relative use of global and local geometric cues. In particular, the Floor Surface-Feature condition appears to have facilitated the use of global geometric cues as evidenced by a greater proportion of geometrically-correct responses in the Rectangle enclosure. The following planned comparisons further investigated the effects of beacons and Surface-Features across enclosure types.

2.3. Planned comparisons

Fig. 5 shows the mean proportion of responses to the geometrically-correct locations for Beacons-Present and Beacons-Absent groups across Testing Enclosures, plotted separately for each Surface-Feature Condition. To more directly investigate the effects of beacons and surface features on the use of global and local geometric cues, separate two-way mixed ANOVAs on mean proportion of responses to the geometrically-correct locations (i.e., top-right and bottom-left) with Enclosure Type (Trapezoid, Rectangle, Parallelogram 1, Parallelogram 2) and Beacon Status (Beacons Present, Beacons Absent) as factors were conducted for each Surface-Feature condition. Additionally, separate one-sample t-tests were conducted to compare each group's performance in each testing enclosure to chance (0.5). Table 2 provides a summary of the obtained results.

2.3.1. Surface feature: None

There was a main effect of Enclosure Type, F(3, 78) = 17.81. p < .001, a significant Enclosure Type × Beacon Status interaction, F(3, 78) = 5.13, p < .01, but no effect of Beacon Status, F(1, 26) = 3.36, p = .08. The effect of Enclosure Type was due to significantly fewer geometrically-correct responses in the Parallelogram 2 enclosure than the other three enclosures, *ps* < .002, and significantly fewer geometrically-correct responses in the Rectangle enclosure than the Parallelogram 1 enclosure, p < .01. The interaction was due to the Beacons-Absent group making significantly more responses in the Parallelogram 1 enclosure, independent-samples t-test, t(26) = 2.15, p = .04, and significantly less in the Parallelogram 2 enclosure, independent-samples *t*-test, t(26) = 3.08, p < .01, to the geometrically-correct locations than the Beacons-Present group. Additionally, the Beacons-Present group made more geometricallycorrect responses than expected by chance in the Trapezoid,



Trapezoid Rectangle Parallel 1 Parallel 2

Testing Enclosures

Trapezoid Rectangle Parallel 1 Parallel 2

Testing Enclosures

0.2

0.0

1.0

0.8

0.6

0.4

0.2

0.0

1.0

0.8

0.6

0.4

0.2

0.0

Prop. Geometric Corr.

Prop. Geometric Corr.

Surface Feature: Floor

Surface Feature: Both

Fig. 5. Panels plot mean proportion of responses to the geometrically-correct locations (i.e., top-right and bottom-left) across testing enclosures for the Beacons-Present (filled) and Beacons-Absent (unfilled) groups separately for each Surface-Feature condition. Dotted borders represent measures that do not differ from chance. Dashed lines represent chance (0.5). Error bars represent standard errors of the means.

Trapezoid Rectangle Parallel 1 Parallel 2

Testing Enclosures

I abic 2

Summary of the obtained allocation of responses to geometrically-correct locations relative to chance performance (0.5) for each group.

	Testing enclosure			
Training condition	Rectangle	Parallelogram 1	Parallelogram 2	Cue type(s) used
Surface Feature: None				
Beacons Present	Above	Above	Equal	Global & Local
Beacons Absent	Equal	Above	Below	Local
Surface Feature: Walls				
Beacons Present	Above	Above	Equal	Global & Local
Beacons Absent	Equal	Above	Equal ^b	Local ^a
Surface Feature: Floor				
Beacons Present	Above	Equal ^c	Equal	Global & Local ^a
Beacons Absent	Above	Above	Equal	Global & Local
Surface Feature: Both				
Beacons Present	Above	Above	Equal	Global & Local
Beacons Absent	Equal	Equal	Equal	-

^a Two of the three predictions held, and the third trended in the predicted direction.

^b Obtained results trended below chance.

^c Obtained results trended above chance.

Rectangle, and Parallelogram 1 enclosures, independent-samples *t*-tests, ts(13) > 2.77, ps < .02, but not in the Parallelogram 2 enclosure independent-samples *t*-test, t(13) = .23, p = .82. In contrast, the Beacons-Absent group made significantly more geometrically-correct responses than expected by chance in the Trapezoid and Parallelogram 1 enclosures, independent-samples *t*-tests, ts(13) > 3.22, ps < .01, and significantly fewer in the Parallelogram 2 enclosure, independent-samples *t*-test, t(13) = 5.3, p < .001. Furthermore, responding did not differ from chance in the Rectangle enclosure, independent-samples *t*-test, t(13) = .59, p = .57. These results suggest that in the absence of surface features and beacons, responding depended only upon local geometric cues. However, the presence of beacons facilitated the use of global geometric cues without hindering the use of local geometric cues.

2.3.2. Surface feature: Walls

There was a main effect of Enclosure Type, F(3, 78) = 9.7, p < .001. However, the effect of Beacon Status, F(1, 26) = 1.6, p = .22, and the Enclosure Type × Beacon Status interaction, F(3, 78) = 1.41, p = .25, were not significant. The effect of Enclosure Type was due to significantly fewer geometrically-correct responses in the Parallelogram 2 enclosure than in the other three enclosures, ps < .01. Additionally, the Beacons-Present group made significantly more geometricallycorrect responses than expected by chance in the Trapezoid, Rectangle, and Parallelogram 1 enclosures, independent-samples *t*-tests, *ts*(13)>2.19, *ps*<.05, but not in the Parallelogram 2 enclosure, independent-samples *t*-test, t(13) = .23, p = .82. In contrast, the Beacons-Absent group made significantly more geometricallycorrect responses than expected by chance in the Parallelogram 1 enclosure, independent-samples t-test, t(13) = 5.78, p < .001, but did not differ from chance in the Trapezoid, Rectangle, or Parallelogram 2 enclosures, independent-samples *t*-tests, ts(13) < 2.01, ps > .06. These results suggest that in the presence of surface features bound to the walls near the goal location, responding depended primarily on local geometric cues. The beacons, however, facilitated the use of global geometric cues without hindering the use of local geometric cues.

2.3.3. Surface feature: Floor

There was a main effect of Enclosure Type, F(3, 78)=5.76, p=.001, a main effect of Beacon Status, F(1, 26)=6.13, p=.02, but the Enclosure Type × Beacon Status interaction was not significant, F(3, 78)=.34, p=.8. The effect of Enclosure Type was due to significantly fewer responses in the Parallelogram 2 enclosure than the other three enclosures, ps < .05. The effect of Beacon Status was due to the Beacons-Absent group making significantly more geometrically-correct responses in the Trapezoid, Rectangle

and Parallelogram 1 enclosures than the Beacons-Present group, independent-samples *t*-tests, ts(13) > 13.65, $ps \le .001$. Additionally, the Beacons-Present group made significantly more geometricallycorrect responses than expected by chance in the Rectangle enclosure, independent-samples t-test, t(13) = 3.18, p < .01, but not in the Trapezoid, Parallelogram 1, and Parallelogram 2 enclosures, independent-samples t-tests, ts(13) < 1.95, ps > .07. In contrast, the Beacons-Absent group made significantly more geometricallycorrect responses than expected by chance in the Trapezoid, Rectangle, and Parallelogram 1 enclosures, independent-samples t-tests, ts(13)>9.81, ps < .001, but not in the Parallelogram 2 enclosure, t(13) = 1.09, p = .3. These results suggest that the line surface feature bound to the floor of the enclosure facilitated the use of global geometric cues, but did not hinder the use of local geometric cues (i.e., at chance in the Parallelogram 2 enclosure). The presence of beacons appears to have reduced the effect of the Floor surface feature without interrupting the use of global and local geometric cues.

2.3.4. Surface feature: Both

There was no effect of Enclosure Type, F(3, 78) = 1.11, p = .35, or Beacon Status, F(1, 26) = 1.95, p = .18, and no interaction, F(3, 78) = .41, p = .74. However, the Beacons-Present group made significantly more geometrically-correct responses than expected by chance in the Trapezoid, Rectangle, and Parallelogram 1 enclosures, ts(13) > 2.34, ps < .05, but not in the Parallelogram 2 enclosure, t(13) = 1.95, p = .07. In contrast, the Beacons-Absent group did not significantly differ from chance in any of the testing enclosures, ts(13) < 1.75, ps > .1. These results suggest that the combination of wall and floor surface features hindered the use of local and global geometric cues (i.e., at chance in the Parallelogram 1 enclosure). The presence of beacons, however, appears to have facilitated the use of global and local geometric cues.

3. Discussion

The presence of the beacons during training facilitated acquisition of the task relative to the absence of the beacons. All Beacons-Present groups acquired the task at an equivalent rate and to an equivalently high level of accuracy, regardless of the presence or absence any surface features. In the absence of beacons, the surface features influenced acquisition of the task. The darkerwalls surface feature facilitated acquisition to the same extent as the presence of beacons. Such a result suggests that surface features tied to the walls of the enclosure at the goal location may have functioned like beacons. However, acquisition in the absence of wall-surface features (i.e., None and Floor) progressed more slowly and the asymptote of the acquisition curves was lower. These results suggest, perhaps not surprisingly, that the task was relatively more difficult to acquire in the absence of disambiguating cues (i.e., beacons, shaded walls). Importantly, all groups reliably selected the correct corner more often than expected by chance for the last three blocks of training.

The distribution of responses in the trapezoid test enclosure (see Fig. 3, right panels), in which the surface features and unique beacons were absent, is suggestive of the relative dependence on global and local geometric cues. For example, participants trained in the absence of surface features and beacons (i.e., Surface Feature: None, Beacons-Absent), allocated the majority of responses to the corners with consistent local-geometric cues (i.e., top-right and bottom-right). In contrast, participants trained in the absence of surface features but with beacons present (i.e., Surface Feature: None, Beacons-Present), allocated the majority of responses to the corners that were consistent with global-geometric cues (i.e., topright and bottom-left). That is, the presence of beacons during training appears to have shifted dependence from local to global cues in the absence of surface features.

The testing enclosures were selected to assess dependence on global-geometric cues in isolation (the Rectangle enclosure), in alignment with local-geometric cues (the Parallelogram 1 enclosure) and in conflict with local-geometric cues (the Parallelogram 2 enclosure). For example (see Fig. 5), participants trained in the absence of surface features and beacons (i.e., Surface Feature: None, Beacons-Absent) responded to the geometrically-correct locations as often as would be expected by chance in the Rectangle enclosure and significantly less than would be expected by chance in the Parallelogram 2 enclosure. That is, consistent with the response distribution in the trapezoid test enclosure, participants' responding appears to depend on local geometric cues. In contrast, participants trained in the absence of surface features but with beacons present (i.e., Surface Feature: None, Beacons-Present) responded to the geometrically-correct locations more often that would be expected by chance in the Rectangle enclosure and as often as would be expected by chance in the Parallelogram 2 enclosure. That is, consistent with the response distribution in the trapezoid test enclosure, the presence of beacons during training appears to have facilitated the use of global cues. These results are consistent with previous research which has shown that a rectangular arrangement of discrete objects is sufficient to produce reliance on global geometric cues (e.g., Gibson et al., 2007; Pecchia and Vallortigara, 2012; Sutton et al., 2012). However, the at-chance performance of the Beacons-Present group in the Parallelogram 2 enclosure suggests that local-geometric cues continued to influence responding. That is, responding depended on both global- and local-geometric cues in the presence of beacons.

Overall, testing results suggest that the presence of beacons and surface features during training differentially influenced the use of global and local geometric cues (refer to Table 2). When beacons were present during training, participants used both global and local geometric cues regardless of the Surface-Feature condition. That is, participants responded to the geometrically-correct location greater than would be expected by chance in the Rectangle enclosure, in which only global geometric cues were available, and they responded at chance in the Parallelogram 2 enclosure, in which global and local geometric cues were misaligned.

When beacons were absent during training, the surface features influenced the use of global and local geometric cues (refer to Table 2). In the absence of any surface features, participants used only the local geometric cues, as evidenced by at-chance responding in the Rectangle enclosure and below-chance responding in the Parallelogram 2 enclosure. In the presence of wall surface features, participants also only used local geometric cues, as evidenced by at-chance responding in the Rectangle enclosure and a trend toward below-chance responding in the Parallelogram 2 enclosure. In the presence of the floor surface feature, participants used global and local geometric cues, as evidenced by above-chance responding in the Rectangle enclosure and at-chance responding in the Parallelogram 2 enclosure. Lastly, in the presence of both the wall and floor surface textures, participants did not use global or local geometric cues, as evidenced by at-chance responding in all test enclosures—including the Parallelogram 1 enclosure in which global and local geometric cues were aligned.

Collectively, it appears that the local geometric cues in the enclosure are used in the absence of any other orientation cues. Surface features applied to the walls of the enclosure appear to facilitate acquisition but not influence the use of local geometric cues, whereas a line applied to the floor which corresponds to the major principal axis of the enclosure appears to facilitate the use of global geometric cues without hindering the use of local geometric cues. The presence of multiple surface features, however, appears to hinder the use of geometric cues. Finally, the presence of beacons in each corner of the enclosure, regardless of any surface features that may be present, appears to facilitate the use of global geometric cues.

The current results are consistent with previous research indicating the use of both global and local geometric cues for reorientation (Bodily et al., 2011; Lubyk et al., 2012) and extend these findings by examining the roles of beacons and surface features, separately and combined, on the use of geometric cues. Specifically, the current results illuminate some subtle nuances regarding the roles of beacons and surface features on the use of global and local and geometric cues. Primarily, they indicate that the influence of beacons and surface features are not functionally equivalent. As result, we believe these results have implications for terminology related to non-geometric cues. Fundamentally, we suggest that, like a distinction between global and local geometric cues, distinctions must be made between surface features and beacon non-geometric cues. Only then will future research be able to delineate the conditions under which various spatial cues are utilized for reorientation.

Acknowledgments

This research was conducted following the relevant ethical guidelines for human research and supported in part by a National Science Foundation Science and Technology Expansion Program (NSF-STEP DUE-0856593) Grant to the College of Science and Technology at AASU, an AASU Summer Research Grant to BRS, and funds from the Office of the Vice President for Research and the Jack N. Averitt College of Graduate Studies at Georgia Southern University to BRS and KDB.

References

- Babb, S.J., Crystal, J.D., 2003. Spatial navigation on the radial maze with trial-unique intramaze cues and restricted extramaze cues. Behav. Processes 64, 103–111.
- Bodily, K.D., Eastman, C.K., Sturz, B.R., 2011. Neither by global nor local cues alone: evidence for a unified orientation process. Anim. Cogn. 14, 665–674.
- Brown, M.F., Bing, M.N., 1997. In the dark: spatial choice when access to spatial cues is restricted. Anim. Learn. Behav. 25, 21–30.
- Cheng, K., 1986. A purely geometric module in the rat's spatial representation. Cognition 23, 149–178.
- Cheng, K., 2008. Whither geometry? Troubles of the geometric module. Trends Cogn. Sci. 12, 355–361.
- Cheng, K., Gallistel, C.R., 2005. Shape parameters explain data from spatial transformations: comment on Pearce et al. (2004) and Tommasi and Polli (2004). J. Exp. Psychol.: Anim. Behav. Processes 31, 254–259.
- Cheng, K., Newcombe, N.S., 2005. Is there a geometric module for spatial orientation? Squaring theory and evidence. Psychon. Bull. Rev. 12, 1–23.
- Cheng, K., Newcombe, N.S., 2006. Geometry, features, and orientation in vertebrate animals: a pictorial review. In: Brown, M.F., Cook, R.G. (Eds.), Anim. Spat. Cogn.: Comp. Neural Comput. Approaches, On-line: www.pigeon.psy.tufts.edu/asc/cheng/

Cole, M.R., Gibson, L., Pollack, A., Yates, L., 2011. Potentiation and overshadowing of shape by wall color in a kite-shaped maze using rats in a foraging task. Learn. Motiv. 42, 99–112.

Gallistel, C.R., 1990. The Organization of Learning. MIT Press, Cambridge, MA.

- Gibson, B.M., Wilks, T.J., Kelly, D.M., 2007. Rats (*Rattus norvegicus*) encode the shape of an array of discrete objects. J. Comp. Psychol. 121, 130–144.
- Healy, S., 1998. Spatial Representation in Animals. Oxford University Press, Oxford, England.
- James, W., 1890. The Principles of Psychology. Holt, New York.
- Krider, R.E., Raghubir, P., Krishna, A., 2001. Pizzas: pi or square? Psychophysical biases in area comparisons. Mark. Sci. 20, 405–425.
- Learmonth, A.E., Nadel, L., Newcombe, N.S., 2002. Children's use of landmarks: implications for modularity theory. Psychol. Sci. 13, 337–341.
- Lubyk, D.M., Dupuis, B., Gutiérrez, L., Spetch, M.L., 2012. Geometric orientation by humans: angles weigh in. Psychon. Bull. Rev. 19, 436–442, http://dx.doi.org/10.3758/s13423-012-0232-z.
- Miller, N.Y., Shettleworth, S.J., 2007. Learning about environmental geometry: an associative model. J. Exp. Psychol.: Anim. Behav. Processes 33, 191–212.
- McGregor, A., Jones, M.J., Good, M.A., Pearce, J.M., 2006. Further evidence that rats rely on local rather than global spatial information to locate a hidden goal: reply to Cheng and Gallistel (2005). J. Exp. Psychol.: Anim. Behav. Processes 32, 314–321.
- Olton, D.S., Collison, C., 1979. Intramaze cues and "odor trails" fail to direct choice behavior on an elevated maze. Anim. Learn. Behav. 7, 221–223.

- Pearce, J.M., Graham, M., Good, M.A., Jones, P.M., McGregor, A., 2006. Potentiation, overshadowing, and blocking of spatial learning based on the shape of the environment. J. Exp. Psychol.: Anim. Behav. Processes 32, 201–214.
- Pecchia, T., Vallortigara, G., 2012. Spatial reorientation by geometry with freestanding objects and extended surfaces: a unifying view. Proc. R. Soc. B: Biol. Sci. 7, 2228–2236, http://dx.doi.org/10.1098/rspb.2011.2522.
- Ratliff, K.R., Newcombe, N.S., 2008. Reorienting when cues conflict: evidence for an adaptive combination view. Psychol. Sci. 19, 1301–1307.
- Sturz, B.R., Bodily, K.D., 2011. Is surface-based orientation influenced by a proportional relationship of shape parameters? Psychon. Bull. Rev. 18, 848–854.
- Sturz, B.R., Forloines, M.R., Bodily, K.D., 2012. Enclosure size and the use of local and global geometric cues for reorientation. Psychon. Bull. Rev., http://dx.doi.org/10.3758/s13423-011-0195-5.
- Sturz, B.R., Gurley, T., Bodily, K.D., 2011. Orientation in trapezoid-shaped enclosures: implications for theoretical accounts of geometry learning. J. Exp. Psychol.: Anim. Behav. Processes 37, 246–253, http://dx.doi.org/10.1037/a0021215.
- Sutton, J.E., Twyman, A.D., Joanisse, M.F., Newcombe, N.S., 2012. Geometry three ways: an fMRI investigation of geometric information processing during reorientation. J. Exp. Psychol.: Learn. Mem. Cogn., http://dx.doi.org/10.1037/a0028456.
- Tommasi, L., Chiandetti, C., Pecchia, T., Sovrano, V.A., Vallortigara, G., 2012. From natural geometry to spatial cognition. Neurosci. Biobehav. Rev. 36, 799–824.

Contents lists available at SciVerse ScienceDirect

Review

Behavioural Processes



journal homepage: www.elsevier.com/locate/behavproc

Abstract concepts: Data from a Grey parrot

Irene M. Pepperberg^{a,b,*}

^a Department of Psychology, Harvard University, United States ^b Department of Psychology, Brandeis University, United States

ARTICLE INFO

Article history: Received 5 July 2012 Received in revised form 13 August 2012 Accepted 4 September 2012

Keywords: Gray parrot cognition Gray parrot numerical concepts Same-different Avian cognition Abstract concepts in gray parrots

ABSTRACT

Do humans and nonhumans share the ability to form abstract concepts? Until the 1960s, many researchers questioned whether avian subjects could form categorical constructs, much less more abstract formulations, including concepts such as same-different or exact understanding of number. Although ethologists argued that nonhumans, including birds, had to have some understanding of divisions such as prey versus predator, mate versus nonmate, food versus nonfood, or basic relational concepts such as more versus less, simply in order to survive, no claims were made that these abilities reflected cognitive processes, and little formal data from psychology laboratories could initially support such claims. Researchers like Anthony Wright, however, succeeded in obtaining such data and inspired many others to pursue these topics, with the eventual result that several avian species are now considered "feathered primates" in terms of cognitive processes. Here I review research on numerical concepts in the Gray parrot (*Psittacus erithacus*), demonstrating that at least one subject, Alex, understood number symbols as abstract representations of real-world collections, in ways comparing favorably to those of apes and young human children. He not only understood such concepts, but also appeared to learn them in ways more similar to humans than to apes.

© 2012 Elsevier B.V. All rights reserved.

Contents

1.	ntroduction	82
2.	Nex's non-numerical capacities	83
3.	Nex's early numerical abilities	83
	3.1. Training and testing methods	84
	3.2. Labeling of basic quantities and simple heterogeneous sets	84
	3.3. Complex heterogeneous sets	84
4.	Nex's more advanced numerical abilities	85
	4.1. Number comprehension	85
	1.2. Use of "none"	85
	1.3. Addition of small quantities	86
	1.4. Ordinality and equivalence	87
	1.5. An exact integer system	88
	4.6. The final study	88
5.	ummary	89
	References	89

1. Introduction

In the early twentieth century, little scientific interest existed in cognitive processes, even in humans. As a consequence, the study

of such processes in nonhumans was also not a viable pursuit. Thus, until the so-called "cognitive revolution" of the 1960s, both ethologists and psychologists, with few exceptions (notably in Europe, e.g., Herz, 1928, 1935; Koehler, 1943), were likely to see nonhumans, and particularly birds, as simple automatons, incapable of complex cognitive processing. Indeed, the term "avian cognition" was considered an oxymoron (see review in Pepperberg, 2011).

Ethologists did accept that birds had to have some understanding of divisions such as prey versus predator, mate versus nonmate,

^{*} Correspondence address: Department of Psychology, Wm James Hall, Harvard University, 33 Kirkland St., Cambridge, MA 02138, United States. Tel : +1 617 495 2852

E-mail addresses: impepper@media.mit.edu, impepper@wjh.harvard.edu

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.09.016

food versus nonfood, or basic relational concepts such as more versus less, simply in order to survive. Ethological research, however, was mostly interested in issues such as "fixed action patterns" (e.g., Tinbergen, 1951)—innate, instinctual behavioral sequences that seemed indivisible and that, once begun, could not be stopped until they ran to completion. Such sequences were initiated by external stimuli known as "releasers," and even removing these releasers mid-stream had no effect. Moreover, because objects that only approximated the releasers might set the behavior in motion, nonhumans were considered incapable of recognizing substitutions or reacting to change of any sort.

Similarly, psychologists concentrated on issues such as stimulus–response chains, where almost all behavior could be explained in terms of histories of positive or negative conditioning to increase or decrease, respectively, behavior toward some external situation. The rules underlying behavior were thought to be the same whatever the species (Skinner, 1938), and species differences were expected to arise only in the speed and extent of acquisition of these rules (for interesting discussions of these ideas see Bitterman, 1965, 1975). The focus was ostensibly on learning, but not in the sense of information processing, remembering, problem solving, rule and concept formation, perception, or recognition: learning was seen as behavior simply being *shaped* by the association of external stimuli and their consequences. Scientists eschewed discussions of issues such as thought, mental representations, or intentional actions (Pepperberg, 1999, 2011).

By the 1960s, however, researchers began to realize that the behavior patterns of their subjects (human or nonhuman) could not be fully explained by current paradigms (e.g., Breland and Breland, 1961). After realizing that even human actions were neither as pre-wired nor as amenable to shaping as once thought, a small group of researchers began to examine nonhumans in the same manner, suggesting a continuum between human and nonhuman abilities (e.g., Hulse et al., 1968). Psychologists such as Herrnstein started examining issues of concept formation in pigeons (e.g., Herrnstein and Loveland, 1964; Herrnstein et al., 1976), and those like Anthony Wright pushed what was then the edge of the envelope to examine so-called "abstract concepts" of same-different (e.g., Premack, 1978); he and his colleagues (Santiago and Wright, 1984; Wright et al., 1984a,b; see also seminal work from the Zentall lab, e.g., Edwards et al., 1983) tried to separate out issues of same-different from those of match-to-sample and nonmatchto-sample and whether subjects were responding on the basis of novelty or other aspects of the task rather than the abstract concept. Specifically, a subject that understands same/different not only knows that two nonidentical blue objects are related in the same way as are two nonidentical green objects-in terms of color-but also knows that the relations between two nonidentical blue objects and two nonidentical square objects are based on the same concept but with respect to a different category, and, moreover, can transfer this understanding to any attribute of an item (Premack, 1978, 1983). Inspired by the research of scientists like Wright and Zentall, my own studies on Gray parrots showed their capacity to understand concepts of category and of same-different (Pepperberg, 1983, 1987a)-and of the absence of same-different (Pepperberg, 1988)-at levels comparable to those of nonhuman primates.

Once Wright, his collaborators, and colleagues helped demonstrate that abstract concept formation was a legitimate area for study in nonhumans, many of us followed their lead to examine other abstract concepts as well. One path that my laboratory took involved the study of a Gray parrot's number concepts. To succeed on number concepts, the bird would have to reorganize how objects were categorized in its world. Specifically, an object would not only be, for example, something to eat or manipulate, or of a specific color or shape, but also would have to be *labeled* with respect to its membership within a quantifiable set, if exact number competence were to be shown. Could a nonhuman acquire that level of abstract understanding? I was hardly the first to study number concepts in nonhumans or even birds, but was the first to examine whether an avian subject could use human number labels symbolically and referentially, to identify exact quantities (see Pepperberg, 2012b). I likely would not have done so had others like Wright not led the way.

Numerical abilities involve many issues. Even for humans, some researchers still disagree on what constitutes various stages of numerical competence; which are the most complex, abstract stages; what mechanisms are involved; and even what is enumerated (for a detailed review, see Carey, 2009). And considerable discussion exists as to the extent to which language-or at least symbolic representation-is required for numerical competence, not only for preverbal children but also for primitive human tribes and nonhumans (e.g., Gordon, 2004; Watanabe and Huber, 2006; Frank et al., 2008). If language and number skills require the same abstract cognitive capacities, then animals, lacking human language and, for the most part, symbolic representation, should not succeed on abstract number tasks; an alternate view is that humans and animals have similar simple, basic number capacities but that only humans' language skills enable development of numerical representation and thus abilities such as verbal counting and addition (see Pepperberg, 2006b; Carey, 2009; Pepperberg and Carey, 2012).

But what if a nonhuman had already acquired a certain level of abstract, symbolic representation? Could such abilities be adapted to the study of numerical competence? Again, with the inspiration and encouragement from colleagues such as Wright, I decided to find out. Here I begin by discussing briefly the background studies with my Gray parrot, Alex, then review the accumulated data that demonstrate the extent of abstract number competence he achieved.

2. Alex's non-numerical capacities

When I first began numerical work with Alex in the 1980s, he had already achieved competence on various tasks once thought limited to young children or at least nonhuman primates (Pepperberg, 1999). Through the use of a modeling technique, roughly based on that of Todt (1975), Alex learned to use English speech sounds to referentially label a large variety of objects and their colors; at the time he could also label two shapes ("3-corner" for triangles, "4-corner" for squares; later he identified various other polygons as "x-corner"). He understood concepts of category: that the same item could be identified with respect to material, color, shape, and object name (e.g., "wood", "blue", "4-corner", and "block"). He had functional use of phrases such as "I want X" and "Wanna go Y", X and Y being appropriate object or location labels. He was acquiring concepts of same, different, and absence-for any object pair he could label the attribute ("color", "shape", and "matter") that was same or different, and state "none" if nothing was same or different; he was also learning to view collections of items and state the attribute of the sole object defined by two other attributes-e.g., in a set of many objects of which some were yellow and some were pentagonal, to label the material of the only one that was both yellow and pentagonal (Pepperberg, 1999). But could he form an entirely new categorical class consisting of quantity labels?

3. Alex's early numerical abilities

As noted above, to succeed on number concepts, Alex would have to reorganize how he categorized objects in his world. He would have to learn that a new set of labels, "one", "two", "three", etc. represented a novel classification strategy; that is, one based on both physical similarity within a group and a group's quantity, rather than solely by physical characteristics of group members. He would also have to generalize this new class of number labels to sets of novel items, items in random arrays, heterogeneous collections, and eventually to more advanced numerical processes (Pepperberg, 1999, 2006b). If successful, he would demonstrate a *symbolic* concept of number, that is, vocally designate the *exact* quantity of a given array with an appropriate numerical, referential utterance in his repertoire (Pepperberg, 2012b).

3.1. Training and testing methods

Via our standard modeling technique that enabled Alex to produce labels for objects, colors, and shapes, he was initially trained to identify small number sets with English labels (note, however, that he initially used "sih" for six, because he had trouble pronouncing the final/s/; Pepperberg, 1987b). As we will see in detail below, this training was quite different from that experienced by young children. For example, unlike children (Carey, 2009), who learn numbers in the appropriate ordinal pattern (i.e., "one", "two", etc.), he was first trained on sets of three and four, because he already had those labels in his repertoire; he was then taught "five" and "two," then "six" and lastly "one." Training in such a manner also ensured that Alex was building his concept of number solely by forming one-to-one associations between specific quantities and their respective number labels (Pepperberg, 2006b). Unlike children, who seem to learn "one" fairly easily (i.e., "one" versus "many", Carey, 2009), "one" was actually rather difficult for Alex to acquire, because he already knew to label a single item with an object label and had to be trained for quite some time to add the number label. Training details are published elsewhere and will not be repeated (Pepperberg, 1987b). Training was limited to sets of a few familiar objects; testing involved transfer to sets of other familiar and novel exemplars. Various publications describe, again in great detail, testing procedures that ensured against myriad forms of possible external cuing, both with respect to inadvertent human cuing and cues based on nonnumber issues such as mass, brightness, density, surface area, odor, item familiarity, or canonical pattern recognition (Pepperberg, 1987b, 1994, 1999, 2006a,b,c).

3.2. Labeling of basic quantities and simple heterogeneous sets

Initial studies demonstrated that Alex could use English labels to quantify small sets of familiar different physical items, up to six, exactly (78.9%, all trials; Pepperberg, 1987b); that is, he overall made few errors, and his data did not show a peak near a correct response with many errors of nearby numerals, which would have suggested only a general sense of quantity (i.e., an approximate number system). Rather, his most common errors across all sets was to provide the label of the object involved-to respond, for example, "key" rather than "four key", which accounted for almost 60% of his roughly 50 errors in \sim 250 trials (another \sim 20% of his errors involved unintelligible responses or misidentifications of the object or material; i.e., 80% of his errors were nonnumerical). Thus Alex indeed had a concept of quantity; he was not, however, necessarily counting, as would a human child who understood, for example, the concept of "five" (Fuson, 1988; Pepperberg, 1999; Mix et al., 2002); that is, who understood the counting principles: that a stable symbolic list of numerals exists, numerals must be applied to individuals in a set to be enumerated in order, they must be applied in 1-1 correspondence, that the last numeral reached in a count represents the cardinal value of the set, and that each numeral is exactly one more than the previous numeral (Gelman and Gallistel, 1986; Fuson, 1988). Even if he was not technically counting, additional tests demonstrated that Alex could quantify even unfamiliar items and those not arranged in any particular (canonical) pattern, such as a square or triangle; he maintained an accuracy of about 75–80% on novel items in random arrays.

Moreover, he could also quantify subsets within heterogeneous sets; that is, in a mixture of X's and Y's, he could respond appropriately to "How many X?", "How many Y?", or "How many toy?" (70%, first trials; Pepperberg, 1987b). Here he outperformed some children, who are generally tested on only homogeneous sets (e.g., Starkey and Cooper, 1995) and who, if asked about subsets within a mixed set of toys, usually label the *total* number of items if, like Alex, they have been taught to label homogeneous sets exclusively (see Siegel, 1982; Greeno et al., 1984).

Despite these tests, we still could not identify the mechanism(s) Alex might be using to succeed. Notably, our tests ensured that Alex could not use nonnumerical cues such as mass, brightness, surface area, odor, object familiarity, or canonical pattern recognition (Pepperberg, 1987b, 1999), because we questioned him on a variety of exemplars of various sizes and of both familiar and novel textures and materials (e.g., metal keys versus bottle corks) often presented by simply tossing them in random arrays on a tray. Such controls did not, however, rule out the possibility that, for the smallest collections, Alex had used a noncounting strategy such as subitizing-a perceptual mechanisms that enables humans to quickly quantify sets up to \sim 4 without counting—or, for larger collections, "clumping" or "chunking"-another form of subitizing (e.g., perception of six as two groups of three; for a review, see von Glasersfeld, 1982). Thus many other tests would be needed to determine the mechanisms that Alex was indeed using.

3.3. Complex heterogeneous sets

Some tests to tease apart subitizing/clumping versus counting issues were initially designed for humans by Trick and Pylyshyn (1989, 1994). In their experiments, subjects had to enumerate a particular set of items embedded within two different types of distractors: (1) white or vertical lines among green horizontals; (2) white vertical lines among green vertical and white horizontals. The authors argued for subitizing for 1-3 in only the first condition, but counting, even for such small quantities, in the second. When subjects thus must distinguish among various items defined by a collection of competing features (e.g., a conjunction of color *and* shape, where an evaluation cannot be made on the basis of a single attribute, such as "whiteness"; see Pepperberg, 1999), subitizing becomes unlikely. Alex could be examined in a comparable manner, because he already was being tested on conjunction (e.g., being asked to identify the color of an item that was both triangular and wood in a collection of differently shaped objects of various materials; Pepperberg, 1992). He could thus be asked to label the quantity of a similarly defined subset-for example, the number of green blocks in a set of orange and green balls and blocks. Would his numerical capacities match those described by Trick and Pylyshyn for humans? (Note that we now understand even more about the effects of the physical dimensions of various stimuli on number competence; see Rugani et al., 2010 for a discussion.)

Alex turned out to be about as accurate as humans (83.3% on 54 trials, Pepperberg, 1994; see Trick and Pylyshyn, 1989), and analyses suggested that he, like humans, was counting. Had he used perceptual strategies similar to those of humans (e.g., subitizing and clumping), rather than counting, he would have made no errors for 1 and 2, few for 3, and more for larger numbers. His errors, however, were random with respect to number of items to be identified (Pepperberg, 1994) and, importantly, his responses were not simply a close approximation to the correct number label (Pepperberg, 1994), which would be expected had he been subitizing or even estimating. In fact, most of Alex's errors seemed unrelated to

numerical competence, but rather were in misinterpreting the defining labels, then correctly quantifying the incorrectly targeted subset. Eight of his nine errors were the correct number for an alternative subset (e.g., the number of blue rather than red keys; in those cases, the quantity of the designated set usually differed from that of the labeled set by two or more items). The problem, however, was that there was no way of knowing whether Alex's perceptual capacities might be more sophisticated than those of humans, allowing him to subitize larger quantities; the data, although impressive with respect to exact number, still did not justify claiming that he was definitively counting. A detailed discussion is in Pepperberg (1994).

In a subsequent study (Pepperberg and Carey, 2012), we further tested Alex's responses concerning exact number. Here we examined how he might process quantities greater than those he could label; we specifically wanted to see if his label "sih" actually referred to exactly six items, or roughly six; that is, to anything he might perceive as large. We showed him, in individual trials under no time constraints, seven, eight, and nine items, asking "How many X?" There were two trials for each quantity, in random order, interspersed with trials on smaller sets and non-number tasks, to ensure that he could switch between sets and objects he could label and those that (potentially) he could not. He was neither rewarded nor scolded whatever his reply, simply told "OK;" we then went to the next query. In trials for sets greater than six, Alex usually initially did not answer, but remained quietly seated on his perch or asked to return to his cage. Only when we continuously badgered him, asking over and over, did he eventually reply "sih." His actions suggested that he knew his standard number answers would be incorrect and he did not, as when was being noncompliant (e.g., see below; Pepperberg, 1992; Pepperberg and Lynn, 2000; Pepperberg and Gordon, 2005), give strings of irrelevant answers, request many treats, or turn his back and preen.

4. Alex's more advanced numerical abilities

En route to determining the mechanism—or mechanisms—Alex used to quantify sets, my students and I examined various other numerical capacities. Thus, Alex was tested on comprehension of numerical labels, on his ability to sum small quantities, and on whether he understood the ordinality of his numbers. The latter task was of particular interest, because, as noted above, unlike children, he had not been trained in an ordinal manner: he had first learned to label sets of three and four, then five and two, then six and one.

4.1. Number comprehension

Although Alex could label numerical sets, he had never been tested on number label comprehension. In general, researchers who teach nonhumans to use a human communication code must ensure the equivalence of label production and comprehension (e.g., Savage-Rumbaugh et al., 1980, 1993), but the issue is particularly important in numerical studies: even a young child who successfully labels the number of items in a small set ("Here's X marbles") might fail when shown a very large quantity and asked "Can you give me X marbles?" That is, the child might not really understand the relationship between the number label and the quantity (Wynn, 1990). If labeling indeed separates animal and human numerical abilities (see above; Watanabe and Huber, 2006; Pepperberg, 2012b; Pepperberg and Carey, 2012), such comprehension-production equivalence would be necessary to demonstrate nonhuman numerical competence (Fuson, 1988).

To test Alex's comprehension abilities, we used a variation of the previous task. Here we simultaneously presented several sets of different quantities of different items—for example, X red blocks,



Fig. 1. Alex's comprehension task. Trials with blocks were the only ones in which all the objects were exactly the same size; these trials tested whether accuracy improved with same-sized objects.

Y yellow blocks, Z green blocks, or X blue keys, Y blue wood, and Z blue pompons, with X, Y, and Z being different quantities. Alex was then queried, respectively, "What color Z?" or "What matter X?" (Pepperberg and Gordon, 2005). He received no training on this task prior to testing. To succeed, he had to comprehend the auditorially presented numeral label (e.g., X = "four") and use its meaning to direct a search for the cardinal amount specified by that label (e.g., four things), that is, know exactly what a set of "X" items is, even when intermixed with other items representing different numerical sets (Fig. 1). We again controlled for contour, mass, etc. by using objects of different sizes, within or across trials so that comprehension of the number label was the only way to perform correctly (Pepperberg and Gordon, 2005). To respond correctly, he also had to identify the item or color of the set specified by the numerical label. Some or all this behavior likely occurred as separate steps, each adding to task complexity (Premack, 1983).

Alex's overall score was again impressive (statistically significant 87.9% on 66 trials), with no errors on the first 10 trials (Pepperberg and Gordon, 2005). Interestingly, errors increased with time, suggesting lack of focus or inattention as testing proceeded. He may have been like keas (Gajdon et al., 2011) or large-billed crows (Izawa and Watanabe, 2011) that will, after succeeding on various tasks, often later employ other, less successful or simply different methods, possibly from boredom (e.g., to engender more interesting responses from trainers; Pepperberg, 2012c) or maybe to find other possible solutions. In any case, he understood the meaning of his number labels somewhat better than young children (see above, Fuson, 1988; Wynn, 1990, 1992), and, most importantly, he had little difficulty with numbers differing by small amounts, suggesting that his number sense was exact and not approximate. Most of his errors appeared to involve color perception or phonological confusion, not numerical misunderstanding: he sometimes erred in distinguishing orange from red or yellow, a consequence of differences in parrot and human color vision (Bowmaker et al., 1994, 1996); he also sometimes confused "wool" and "wood", or "truck" and "chalk"; he pronounced the last label a bit like "chuck" (Pepperberg and Gordon, 2005).

4.2. Use of "none"

The comprehension study was notable for another reason: Alex's spontaneous transfer of use of "none"—learned as a response to the queries "What's same/different?" with respect to two objects when no category (color, shape, or material) was same or different (Pepperberg, 1988)—to the absence of a set of a particular quantity. After responding appropriately for several trials of the standard comprehension task, Alex reacted in a manner quite different from the norm. When, on one particular trial, he was asked "What color three?" to a set of two, three, and six objects, he replied "five"; obviously no such set existed and his response made little sense. The questioner asked twice more, each time he replied "five." He was obviously refusing to answer the question posed, but, unlike his usual responses when he was being noncompliant (e.g., Pepperberg and Lynn, 2000)-that is, when he refused to maintain his gaze on the tray, but instead endlessly preened, made requests to be returned to his cage or for treats he then discarded, or uttered strings of irrelevant labels (e.g., colors not on the tray and thus not possible response choices), here he was providing the label of a number that was not being tested and consistently repeating it. The response seemed irrelevant, but was different enough from noncompliance that, finally, the questioner said "OK, Alex, tell me, what color 5?", to which he immediately responded "none" (Pepperberg and Gordon, 2005). The response came as a complete surprise, as he had never been taught the concept of absence of quantity nor to respond to absence of an exemplar. He had, previously, spontaneously transferred use of "none" from the same-different study to "What color bigger?" for two equally sized items in a study on relative size (Pepperberg and Brezinsky, 1991), but that use of "none" still referred to the absence of difference in an attribute. "None," or a zero-like concept, is advanced, abstract, and relies on the violation of an expectation of presence (Bloom, 1970; Hearst, 1984; Pepperberg, 1988). Of additional interest was that Alex not only had provided a correct, novel response, but had also manipulated the trainer into asking the question he apparently wished to answer, which suggested other levels of abstract processing (Pepperberg and Gordon, 2005). Alex also correctly answered additional queries about absent sets of one to six items, showing that his behavior was not a chance response.

A subsequent study (Pepperberg and Carey, 2012) further emphasized Alex's number comprehension and made use of his knowledge of absence. Here we again tested him with sets larger than he could label: he saw four trays with sets of various numbers of items, including 7 or 8 but omitting 6 (e.g., 3 yellow wool, 4 blue wool, 7 green wool), and was asked "What color six?" to see if he would reply "none" (Pepperberg and Gordon, 2005)-would he require exactly six or accept the set that was roughly six (here, say "green")? These questions tested whether he knew that "six" meant exactly six and not approximately six, that is, whether he truly understood that his labels referred to very specific sets. He was also asked about an existing set for two arrays to ensure he did not learn to respond "none." Thus he had six queries: two probing an existing set (one for a 3-item set, one for 5) and four for which the correct response was "none" if "six" meant exactly six. Alex responded "none" on all four trials involving quantities above six. On trials for colors of sets that were present, he gave the appropriate labels (respectively, "yellow" and "green" to 3- and 5-item sets).

A critical issue was that Alex's initial use of "none" was spontaneous, unlike that of the chimpanzee, Ai, who had to be trained to use the label "zero" (Biro and Matsuzawa, 2001). But our data did not demonstrate whether he really understood the overall *concept* of zero. How similar was his understanding to that of a young child or an adult human? Only additional studies could provide that information.

4.3. Addition of small quantities

Although I had always wanted to determine if Alex could perform the same kind of small number addition as did chimpanzees (Boysen and Berntson, 1989), I had started to focus on other areas of cognitive processing (e.g., research on optical illusions, Pepperberg et al., 2008) at this time. Thus studies on addition (Pepperberg, 2006a), like those on "zero," were unplanned, and came about as follows. Alex, who routinely interrupted the sessions of a younger parrot, Griffin, with phrases like "Talk clearly" or with an appropriate answer, appeared to sum the clicks over the individual trials that we were using to train Griffin on sequential auditory numbers (training to respond to, e.g., three computer-generated clicks with the vocal label "three"). Given how difficult it would be to demonstrate true summation auditorially, I chose to replicate, as closely as possible, the object-based addition study of Boysen and Berntson (1989) on apes, and to use the experiment to study further Alex's understanding of zero (Pepperberg, 2006a).

I chose the Boysen and Bernston procedure because it was a formal test of addition-having a subject observe two (or more) separate quantities and provide the exact label for their total (Dehaene, 1997)-that is, it required both summation and symbolical labeling of the sum by a nonhuman. Most additive and subtractive studies on nonhumans required the subject to choose the larger amount of two sets, not label final quantity (review in Pepperberg, 2006a). Specifically, when the correct response involves choice of relative amount, no information is obtained on whether the subject has "...a digital or discrete representation of numbers" (Dehaene, 1997, p. 27; see also Carey, 2009, for a discussion of how such responses can rely on an approximate number system). In contrast to most other addition studies, moreover, I avoided use of only one token type of a standard size (e.g., whole marshmallows), which could allow evaluations to be based on contour and mass, not number (note Mix et al., 2002).

The procedure was as follows: Alex was presented with a tray on which two upside down cups had been placed (Fig. 2); prior to presentation, a trainer had hidden items such as randomly shaped nuts, bits of cracker, or differently sized jelly beans under each cup, with items in the same cup less than 1 cm from each other. We occasionally used identical candy hearts to see if accuracy was higher when mass/contour cues were available (Pepperberg, 2006a). After bringing the tray up to Alex's face, the experimenter lifted the cup on his left, showed what was under the cup for 2–3 s in initial trials, replaced the cup over the quantity; then replicated the procedure for the cup on his right. For reasons described below, in trials comprising the last third of the experiment, Alex had ~6-10 s to view items under each cup before everything was covered. The experimenter then made eye contact with Alex, who was asked, vocally, and without any training, to respond to "How many total?" He saw collections with all possible addends, totaling to every amount from 1 to 6, plus trials with nothing under both cups to see if he would generalize use of "none" without instruction. No objects other than the cups were visible during questioning. To respond correctly, Alex had to remember the quantity under each cup, perform some combinatorial process, then produce a label for the total amount. He had no time limit in which to respond, given that his



Fig. 2. Alex's addition task.

response time generally correlated with his current interest in the items being used in the task, rather than the task itself (Pepperberg, 1988). Appropriate controls for cuing and tests for interobserver agreement were, as usual, in place (Pepperberg, 2006a).

For sets of countable objects, Alex had a statistically significant accuracy of 85.4% on 48 first trial responses (Pepperberg, 2006a), and his accuracy did not improve on trials with identical tokens. He had trouble with one set of trials, however. Interestingly, when given only 2–3 s, he always erred on the 5 + 0 sum, consistently stating "6"; when given \sim 6–10 s, however, his accuracy went to 100%. Differences in accuracy between the shorter and longer interval trials was significant only on 5 + 0 trials (Fisher's exact test, *p* = 0.01). Such data suggest that he used a counting strategy for 5: only when beyond the subitizing range of 4 did he, like humans, need time to label the set exactly (for a detailed discussion, see Pepperberg, 2006a). Overall, his data are comparable to those of young children (Mix et al., 2002) and, because he added to six, are more advanced than those published on apes (Boysen and Hallberg, 2000).

In a subsequent study (Pepperberg, 2012a), Alex showed that he could perform with equal accuracy when asked to sum three sets of sequentially presented objects-that is, collections of variously sized objects now hidden under three cups. Here he had to maintain numerical accuracy under what could be an additional memory load, because the protocol required two updates in memory rather than one. His first trial score was 8/10 correct, 80%, p < 0.001 (binomial test, chance of either ¹/₄ or 1/6; 1/6 represented a guess of all possible number labels, 1/4 represented a guess of using one of the three addends as well as their sum). For all trials, his score was 10/12 correct or 83.3%. Occasionally, one cup contained no objects, but even if only those trials are considered in which all three cups contained items, Alex's first trial score was 4/5 correct, p = 0.015 (chance of $\frac{1}{4}$; for chance of $\frac{1}{6}$, p < 0.01); his all trials score was 5/6 or, again, 83.3%. In this three-cup task, all of the addends were within subitizing range (Boysen and Hallberg, 2000; Pepperberg, 2006a); thus Alex could easily have tracked these without specifically counting. However, he still would have needed to remember the values under each of the three cups, for several seconds for each cup, and update his memory after seeing what was under each cup, even if nothing was present. Again, because he added up to 6, his competence surpassed that of an ape similarly tested (Sheba: Boysen and Hallberg, 2000).

Interestingly, in the two-cup task, Alex did not respond "none" when nothing was under any cup (Pepperberg, 2006a; NB: such trials were not present in the three-cup task). He either looked at the tray and said nothing (five trials) or said "one" (three trials). He never said "two," showing that he understood that the query did not correspond to the number of cups. On trials in which he did not respond, his lack of action suggested that he knew his standard number answers would be incorrect. Again, he did not react as he did when noncompliant (see above, Pepperberg and Lynn, 2000). His behavior somewhat resembled that of autistic children (Diane Sherman, personal communication, 2005), who simply stare at the questioner when asked "How many X?" if nothing exists to count. As for his response of "one," he may, despite never having been trained on ordinality and having learned numbers in random order (see above), have inferred that "none" and "one" represented the lower end of the number spectrum and conflated the two labels. Such confusion was demonstrated by the chimpanzee Ai (Biro and Matsuzawa, 2001). Alex's inability to use "none" here might have arisen because he was asked to denote the total absence of labeled objects; previously, he was responding to the absence of an attribute. Specifically, these data confirmed that Alex's use of "none" was merely zero-like: he did not use "none", as he did his number labels, to denote a specific numerosity (Pepperberg, 1987b). In that sense, he was like humans in earlier cultures, or young children, who seem to have to be \sim 4 years old before achieving full adult-like understanding of the label for zero (e.g., Wellman and Miller, 1986; Bialystok and Codd, 2000).

4.4. Ordinality and equivalence

As noted above, despite having learned his number labels out of order-quite unlike children-Alex may have deduced something about ordinality, that is, about an exact number line. He had a concept of bigger and smaller (Pepperberg and Brezinsky, 1991) and, without explicit training, may have organized his number labels in that manner. Such behavior would be important for two reasons. First, even for apes that referentially used Arabic symbols, ordinality did not emerge but had to be trained (e.g., Matsuzawa et al., 1991; Boysen et al., 1993; Biro and Matsuzawa, 2001); if Alex understood ordinality without training, his concepts would be more advanced than those of a nonhuman primate. Second, ordinality is intrinsic to verbal counting (e.g., Gelman and Gallistel, 1986; Fuson, 1988). To count, an organism must produce a standard sequence of number tags and know the relationships among and between these tags; for example, that "three" (be it any vocal or physical symbol) not only comes before "four" in the verbal sequence but also represents a quantity less than "four." An understanding of ordinality, therefore, would help support our possible claims for counting.

Notably, ordinality is not a simple concept. Children acquire ordinal–cardinal abilities in steps. They learn cardinality, slowly, usually over the course of over a year, for very small numbers (<4) and a general sense of "more versus less" while acquiring a meaningless, rote ordinal number series. Only around the time that they acquire an understanding of "fourness" do they connect their knowledge of quantity in the small sets with this number sequence to form 1:1 correspondences that can be extended to larger amounts for both cardinal and ordinal accuracy (e.g., Carey, 2009; see also Mix et al., 2002). Children may give the impression that they have full understanding of cardinality before they actually do, by learning associative rules (i.e., respond correctly to "How many?" but fail on "Give me X"; see above) but cannot act in that manner with ordinality (e.g., Teubal and Guberman, 2002; Bruce and Threfall, 2004).

To test what Alex might know about ordinality and compare his abilities to those of children would require first that he learn to label Arabic numerals, so that he could be tested abstractly; that is, in the absence of physical sets of objects. If, after learning English labels for Arabic numerals (production and comprehension) in the absence of the physical quantities to which they refer, Alex could-without any training-use the commonality of these English labels to equate quantities (sets of physical objects) and Arabic numerals, then I could use a task involving these equivalence relations (Pepperberg, 2006c): I could ask him which of two Arabic numerals was bigger or smaller. To ensure that I could repeat the trials enough times to gain statistical significance without Alex learning rote responses to specific pairs, the task would be to identify the color of one of a pair of Arabic numbers (e.g., a green 2, a yellow 5, next to each other on a tray; Fig. 3) that was numerically (not physically) bigger or smaller. He already answered "What color/matter bigger/smaller?" for object pairs and responded "none" for same-sized pairs (Pepperberg and Brezinsky, 1991). To succeed on this new task, he would have to use deductions and inferences: deduce that an Arabic symbol has the same numerical value as its vocal label, compare representations of quantity for which the labels stand, infer rank ordering based on these representations, then state the result orally (Pepperberg, 2006c). Unlike the tasks used in other nonhuman studies (e.g., Olthof et al., 1997; Olthof and Roberts, 2000), the question would not always be about the larger set, and specific stimuli within pairs would not be associated with reward of the corresponding number of items.



Fig. 3. Alex's ordinality task.

To ensure that Alex really did understand not only ordinality but also the meaning of the Arabic numerals, he was tested on several related tasks (Pepperberg, 2006c). Trials on identical numerals of different colors but of the same size (e.g., 6:6) tested if Alex would, as expected, reply "none" to the query as to which was bigger or smaller. To determine if he might be tricked into responding based on the physical appearance of the numerals rather than their meanings, he was queried about numerals of the same value but different colors and different sizes (e.g., 2:2). By mixing Arabic symbols and physical items, I could determine whether he really did understand that, for example, one numeral (an Arabic 6) was bigger than five items (or an Arabic 2 as the same as two items) and cleanly separate mass and number.

Alex did indeed succeed on the equivalence task and, as a consequence, demonstrated that, without direct, explicit training, he inferred the ordinality of his number labels (Pepperberg, 2006c). Notably, he had never been trained to recite the labels in order nor to associate any Arabic numeral with any specific set of objects. Nevertheless, for trials on two different Arabic numbers of the same physical size, his first trial score was 63/84, or 75% (p < 0.01, binomial test, chance of 1/2). If his occasional responses of the Arabic number label rather than the requested color (technically correct, but not with respect to the actual query) were not counted as errors, his score was 74/84, or 88.1% (p < 0.001, binomial test, chance of $\frac{1}{2}$). As in previous studies, errors sometimes involved yellow-orangered confounds. When numerals were the same value-same size, his accuracy was 10/12, or 83.3%, p < 0.01 (binomial test, chance of 1/3; answers could be one of the two colors or "none"). Importantly, statistical comparisons on his first and final trials for all these sets showed no significant differences in accuracy, suggesting that no training was occurring. For the same value-different size trials, counting as correct either "none" or the color label of the physically targeted number, his accuracy was 12/12, or 100%, p < 0.01 (binomial test, chance of 2/3, a color or "none"). Seven times he gave the correct color of the physically targeted number, five times he said "none," but gave colors most often in earlier trials and "none" most often in later trials, as if he shifted after experience with responses based on symbolic value, even though he had initially been rewarded for responses based on physical size (Pepperberg, 2006c).

Alex's responses to trials that mixed objects and numerals were intriguing. For arrays in which object sets were paired with a single Arabic number representing a quantity larger than or equal to the array (incongruent trials) and in which the single Arabic number represented a quantity less than the array (congruent trials), his accuracy was 16/21, or 76.2%, p < 0.01. However, in five trials in which a *single* object was paired with a *single* Arabic number

that represented a larger quantity, Alex consistently replied "none." Only here did the physical set consistently overwhelm symbolic responses.

Overall, Alex did appear 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. Once children learn ordinality and the successor function—that each digit in their number line is one more than the previous digit—they no longer need to be taught the values of each individual digit for digits greater than 4 (Carey, 2009). Why was Alex unlike children in this instance? Might the issue be Alex's difficulty in learning to produce the English sounds? In order to produce any given English label, Alex had to learn to coordinate his syrinx, tracheal muscles, glottis, larynx, tongue height and protrusion, beak opening, and even esophagus (Patterson and Pepperberg, 1998); might there be a way to dissociate vocal and conceptual learning to test this possibility?

4.5. An exact integer system

To test whether such a dissociation existed, colleagues and I devised the following experiment. Initially, I would teach 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. Only after the labels were being produced clearly would I train him to understand that 6<7<8, that is, where the new numerals fit on the number line. He could then be tested as to whether he understood the relationships among 7 and 8 and his other Arabic labels. If he inferred the new number line in its totality, he could be tested on whether, like children, he could 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 (Pepperberg and Carey, 2012). Nothing in his training at this point would provide specific information about the value of 7 and 8; they could refer to ten and twenty items, respectively. The question was whether, all other numerals having been taught as either +1 or -1 than those he already knew (that is, after learning "3" and "4", he was taught "5" and "2", then "6" and "1", Pepperberg, 1987b, 1994), he could use past and present information to induce the cardinal meaning of the labels "seven" and "eight" from their ordinal positions on an implicit count list.

Over the course of the study, Alex did indeed learn to label the novel Arabic numerals, to place them appropriately in his inferred number line, and to label appropriately, on first trials, novel sets of seven and eight physical items. Detailed data is presented in the published paper (Pepperberg and Carey, 2012); the conclusion was that Alex, like children, and unlike nonhuman primates tested so far, 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.

4.6. The final study

Once Alex had acquired the numerals through 8, we went back to the addition task to determine if he could, like apes (Boysen and Berntson, 1989) sum the Arabic numerals that had been hidden under cups (Pepperberg, 2012a). Such a task would demonstrate further knowledge of the representational nature of the numerals. As in the addition experiment with sets of items, he was sequentially shown two Arabic numerals initially hidden under cups and, in their subsequent 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, and asked for the color of the numeral representing the sum; colors 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"). Alex's passing precluded completion of this latter task, but had he lived longer, this procedure, with its additional step, would have allowed testing the same sums many more times without training him to produce a specific response, unlike tasks given other nonhuman subjects (see discussion in Pepperberg, 2012a).

For the Arabic numeral task requiring a numerical response, Alex demonstrated some competence in summing two Arabic numerals, each representing quantities less than or equal to 5, to a total of 1–8. His first trial score was 9/12 (75%), p = 0.004 (chance of 1/3; p = 0.001 for chance of 1/8). His all trials score was 12/15 (80%). Although the study did not contain enough trials to test all possible sums and combinations of addends or to repeat most queries, Alex was given at least one trial for each sum from 1 to 8. The lack of replication of the various sums over trials, however, emphasizes the first trial nature of the results and shows that no training could have been involved. Notably, if the numerals had only approximate meanings, Alex's errors would likely have exhibited a range close to the correct response. In contrast, such was the case only once (Pepperberg, 2012a); the other errors were to state "eight" when the sums were five and four. He thus seemed to have some fixation on producing the label "eight," which was his newest. 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 (see Carey, 2009, for a review).

Because of his death, he had only three trials on queries requiring a color response; his first trial score, 2/3 (66%), was too low for statistical significance (p = 0.07), but the small number of trials preclude real statistical power. His all trials score was 3/4 (75%). These data do, however, suggest a capacity for exact number representation: conceivably, his one error, on the first trial, may have represented a misunderstanding of the task. His response, which labeled the numeral representing "two," suggests he might have responded to the number of objects under the cups (i.e., the two numerals) rather than their values, given that no training of any sort had preceded questioning on this novel task, and all previous queries did refer to the number of objects. Note, however, that he did not persist in this response but was correct when asked a second time and responded appropriately on the next two trials. Overall, Alex, like Sheba, had had no training on summing the Arabic numerals, and, like Sheba, spontaneously transferred from summing items to summing symbols. His data on the color response task (although extremely limited)-a task somewhat like that of Sheba's, in that possible responses were available from which to choose-tended toward significance. In contrast to Sheba, however, he had to indicate the label not just for the sum but also 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.

5. Summary

The above data demonstrate the extent to which a nonhuman, nonprimate, nonmammalian subject can form complex, abstract concepts and, specifically, that one particular subject, Alex, understood the cardinal representation of his vocal number labels and their corresponding Arabic numerals. He succeeded at levels that, on occasion, went beyond those of nonhuman primates and approached those of children.

Other nonhumans have, of course, succeeded on related numerical tasks, but none yet, like Alex, have deduced the successor function. For example, nonhumans can represent ordinal relations

among arbitrary stimuli, even among Arabic digits, but without necessarily having knowledge of these symbols' cardinal values (e.g., Emmerton et al., 1997; Harris et al., 2007; Beran et al., 2008; Matsuzawa, 2009). Nonhumans also use an analog magnitude system to evaluate more/less for sets of items and transfer to new set sizes, but their data are constrained by Weber's Law (e.g., Brannon and Terrace, 1998, 2000; Emmerton and Renner, 2009; Scarf et al., 2011; note Dehaene, 2009); that is, their evaluations are not precise but center around the correct response, unlike Alex's results. Similarly, some nonhumans engage in approximate addition and subtraction (e.g., Rugani et al., 2009), and have even mapped numerals to approximate quantities, but these latter results could be related to hedonic value or reward probablility (e.g., Beran et al., 2008; note Olthof et al., 1997). Only those nonhumans that symbolically map numerals to exact cardinal values of sets (notably, Matsuzawa's Ai, Boyen's Sheba, and Alex), seem able to engage in several types of precise numerical computations, especially for quantities above 4.

Finally, Alex was not the only avian subject to succeed on many cognitive tasks. Other Gray parrots have succeeded on tasks involving exclusion (e.g., Mikolasch et al., 2011), corvids are considered "feathered primates" (e.g., Emery and Clayton, 2004), Vallortigara and colleagues have shown advanced abilities in chicks (Vallortigara, 2012) and Wright's work and those of his colleagues (see many papers in this special edition) demonstrate advanced cognitive abilities in pigeons. We have come a long way from the 1960s, and much of our progress was inspired by Anthony Wright, who was among the first to challenge the status quo and argue that nonhumans should be tested for the same types of abstract concept formation as humans.

References

- Beran, M.J., Harris, E.H., Evans, T.A., Klein, E.D., Chan, B., Flemming, T.M., et al., 2008. Ordinal judgements of symbolic stimuli by capuchin monkeys (*Cebus paella*) and rhesus monkeys (*Macaca mulatta*): the effects of differential and nondifferential reward. J. Comp. Psychol. 122, 52–61.
- Bialystok, E., Codd, J., 2000. Representing quantity beyond whole numbers: some, none and part. Can. J. Exp. Psychol. 54, 117–128.
- Biro, D., Matsuzawa, T., 2001. Use of numerical symbols by the chimpanzee (*Pan troglodytes*): cardinals, ordinals, and the introduction of zero. Anim. Cogn. 4, 193–199.
- Bitterman, M.E., 1965. The evolution of intelligence. Sci. Am. 212, 92-100.
- Bitterman, M.E., 1975. The comparative analysis of learning. Science 188, 699-709.
- Bloom, L., 1970. Language Development: Form and Function in Emerging Grammars. MIT Press, Cambridge, MA.
- Bowmaker, J.K., Heath, L.A., Das, D., Hunt, D.M., 1994. Spectral sensitivity and opsin structure of avian rod and cone visual pigments. Invest. Ophthalmol. Vis. Sci. 35, 1708.
- Bowmaker, J.K., Heath, L.A., Wilkie, S.E., Das, D., Hunt, D.M., 1996. Middlewave cone and rod visual pigments in birds: spectral sensitivity and opsin structure. Invest. Ophthalmol. Vis. Sci. 37, S804.
- Boysen, S.T., Berntson, G.G., 1989. Numerical competence in a chimpanzee (*Pan troglodytes*). J. Comp. Psychol. 103, 23–31.
- Boysen, S.T., Hallberg, K.I., 2000. Primate numerical competence: contributions toward understanding nonhuman cognition. Cogn. Sci. 24, 423–443.
- Boysen, S.T., Berntson, G.G., Shreyer, T.A., Quigley, K.S., 1993. Processing of ordinality and transitivity by chimpanzees (*Pan troglodytes*). J. Comp. Psychol. 107, 208–215.
- Brannon, E.M., Terrace, H.S., 1998. Ordering of the numerosities 1–9 by monkeys. Science 282, 746–749.
- Brannon, E.M., Terrace, H.S., 2000. Representation of the numerosities 1–9 by rhesus macaques. J. Exp. Psychol. Anim. Behav. Process 26, 31–49.
- Breland, K., Breland, M., 1961. The misbehavior of organisms. Am. Psychol. 16, 661–664.
- Bruce, B., Threfall, J., 2004. One, two, three and counting. Educ. Stud. Math. 55, 3-26.
 - Carey, S., 2009. The Origin of Concepts. Oxford University Press, New York.
 - Dehaene, S., 1997. The Number Sense. Oxford University Press, Oxford, UK.
 - Dehaene, S., 2009. Origins of mathematical intuitions: the case of arithmetic. Ann. N.Y. Acad. Sci. 1156, 232–259.
 - Edwards, C.A., Jagielo, J.A., Zentall, T.R., 1983. Same/different symbol use by pigeons. Anim. Learn. Behav. 11, 349–355.
- Emery, N.J., Clayton, N.S., 2004. The mentality of crows: convergent evolution of intelligence in corvids and apes. Science 306, 1903–1907.
- Emmerton, J., Lohmann, A., Niemann, J., 1997. Pigeons' serial ordering of numerosity with visual arrays. Anim. Learn. Behav. 25, 234–244.

- Emmerton, J., Renner, J.C., 2009. Local rather than global processing of visual arrays in numerosity discrimination by pigeons (*Columba livia*). Anim. Cogn. 12, 511–526.
- Frank, M.C., Everett, D.L., Fedorenko, E., Gibson, E., 2008. Number as a cognitive technology: evidence from Pirahã language and cognition. Cognition 108, 819–824.
- Fuson, K.C., 1988. Children's Counting and Concepts of Number. Springer-Verlag, Berlin, Heidelberg, New York.
- Gajdon, G.K., Amann, L., Huber, L., 2011. Keas rely on social information in a tool use task but abandon it in favor of overt exploration. Interact. Stud. 12, 304–323. Gelman, R., Gallistel, C.R., 1986. The Child's Understanding of Number, 2nd ed. Har-
- vard University Press, Cambridge, MA.
- Gordon, P., 2004. Numerical cognition without words: evidence from Amazonia. Science 306, 496–499.
- Greeno, J.G., Riley, M.S., Gelman, R., 1984. Conceptual competence and children's counting. Cogn. Psychol. 16, 94–143.
- Harris, E.H., Beran, M.J., Washburn, D.A., 2007. Ordinal-list integration for symbolic, arbitrary, and analog stimuli by rhesus macaques (*Macaca mulatta*). J. Gen. Psychol. 134, 183–197.
- Hearst, E., 1984. Absence as information: some implications for learning, performance and representational processes. In: Roitblat, H.L., Bever, T.G., Terrace, H.S. (Eds.), Animal Cognition. Erlbaum, Hillsdale, NJ, pp. 311–332.
- Herrnstein, R.J., Loveland, D.H., 1964. Complex visual concept in the pigeon. Science 146, 549–551.
- Herrnstein, R.J., Loveland, D.H., Cable, C., 1976. Natural concepts in pigeons. J. Exp. Psychol. Anim. Behav. Process 2, 285–311.
- Herz, M., 1928. Wahrnehmungpsychologische untersuchungen am eichelhäher. Z. Vergleich. Physiol. 7, 144–194.
- Herz, M., 1935. Die untersuchungen über den formensinn der honigbiene. Naturwiss 23, 618–624.
- Hulse, S.H., Fowler, H., Honig, W.K. (Eds.), 1968. Cognitive Processes in Animal Behavior. Erlbaum, Hillsdale, NJ.
- Izawa, I.-E., Watanabe, S., 2011. Observational learning in the large-billed crow (*Corvus macrorhynchos*): effect of demonstrator-observer dominance relationship. Interact. Stud. 12, 281–303.
- Koehler, O., 1943. 'Zähl'-Versuche an einem kolkraben und vergleichsversuche an menschen. Z. Tierpsychol. 5, 575–712.
- Matsuzawa, T., 2009. Symbolic representation of number in chimpanzees. Curr. Opin. Neurobiol. 19, 92–98.
- Matsuzawa, T., Itakura, S., Tomonaga, M., 1991. Use of numbers by a chimpanzee: a further study. In: Ehara, A., Kimura, T., Takenaka, O., Iwamoto, M. (Eds.), Primatology Today. Elsevier, Amsterdam, pp. 317–320.
- Mikolasch, S., Kotrschal, K., Schloegl, C., 2011. African grey parrots (*Psittacus erithacus*) use inference by exclusion to find hidden food. Biol. Lett. 7, 875–877.
- Mix, K., Huttenlocher, J., Levine, S.C., 2002. Quantitative Development in Infancy and Early Childhood. Oxford University Press, New York.
- Olthof, A., Roberts, W.A., 2000. Summation of symbols by pigeons (*Columba livia*): the importance of number and mass of reward items. J. Comp. Psychol. 114, 158–166.
- Olthof, A., Iden, C.M., Roberts, W.A., 1997. Judgments of ordinality and summation of number symbols by squirrel monkeys (*Saimiri sciureus*). J. Exp. Psychol. Anim. Behav. Process. 23, 325–333.
- Patterson, D.K., Pepperberg, I.M., 1998. A comparative study of human and Grey parrot phonation: acoustic and articulatory correlates of stop consonants. J. Acoust. Soc. Am. 103, 2197–2213.
- Pepperberg, I.M., 1983. Cognition in the African Grey parrot: preliminary evidence for auditory/vocal comprehension of the class concept. Anim. Learn. Behav. 11, 179–185.
- Pepperberg, I.M., 1987a. Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): learning with respect to categories of color, shape, and material. Anim. Learn. Behav. 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., 1988. Acquisition of the concept of absence by an African Grey parrot: learning with respect to questions of same/different. J. Exp. Anal. Behav. 50, 553–564.
- Pepperberg, I.M., 1992. Proficient performance of a conjunctive, recursive task by an African Grey parrot (*Psittacus erithacus*). J. Comp. Psychol. 106, 295–305.
- Pepperberg, I.M., 1994. Evidence for numerical competence in an African Grey parrot (*Psittacus erithacus*). J. Comp. Psychol. 108, 36–44.
- Pepperberg, I.M., 1999. The Alex Studies. Harvard University Press, Cambridge, MA. Pepperberg, I.M., 2006a. Grey Parrot (*Psittacus erithacus*) numerical abilities: addi-
- tion and further experiments on a zero-like concept. J. Comp. Psychol. 120, 1–11.
- Pepperberg, I.M., 2006b. Grey parrot numerical competence: a review. Anim. Cogn. 9, 377–391.
- Pepperberg, I.M., 2006c. Ordinality and inferential abilities of a Grey Parrot (*Psittacus erithacus*). J. Comp. Psychol. 120, 205–216.

- Pepperberg, I.M., 2011. Avian cognition and social interaction: fifty years of advances. Interact. Stud. 12, 195–207.
- Pepperberg, I.M., 2012a. Further evidence for addition and numerical competence by a Grey parrot (*Psittacus erithacus*). Anim. Cogn. 15, 711–717.
- Pepperberg, I.M., 2012b. Symbolic communication in the Grey parrot. In: Vonk, J., Shackelford, T. (Eds.), Oxford Handbook of Comparative Evolutionary Psychology. Oxford University Press, New York (Chapter 16).
- Pepperberg, I.M., 2012c. Emotional birds—or advanced cognitive processing? In: Watanabe, S., Kuczaj, S. (Eds.), Emotions of Animals and Humans: Comparative Perspectives. Springer, Japan, (Chapter 3).
- Pepperberg, I.M., Brezinsky, M.V., 1991. Acquisition of a relative class concept by an African Grey Parrot (*Psittacus erithacus*): discriminations based on relative size. J. Comp. Psychol. 105, 286–294.
- Pepperberg, I.M., Carey, S., 2012. Grey Parrot number acquisition: the inference of cardinal value from ordinal position on the numeral list. Cognition, http://dx.doi.org/10.1016/j.cognition.2012.07.003.
- Pepperberg, I.M., Gordon, J.D., 2005. Number comprehension by a Grey parrot (*Psit-tacus erithacus*), including a zero-like concept. J. Comp. Psychol. 119, 197–209.
- Pepperberg, I.M., Lynn, S.K., 2000. Perceptual consciousness in Grey parrots. Am. Zool. 40, 393-401.
- Pepperberg, I.M., Vicinay, J., Cavanagh, P., 2008. The Müller–Lyer illusion is processed by a Grey Parrot (*Psittacus erithacus*). Perception 37, 765–781.
- Premack, D., 1978. On the abstractness of human concepts: why it would be difficult to talk to a pigeon. In: Hulse, S.H., Fowler, H., Honig, W.K. (Eds.), Cognitive Processes in Animal Behavior. Erlbaum, Hillsdale, NJ, pp. 421–451.
- Premack, D., 1983. The codes of man and beast. Behav. Brain Sci. 6, 125–176.
- Rugani, R., Regolin, L., Vallortigara, G., 2009. Rudimental numerical competence in 5-day-old domestic chicks (*Gallus gallus*): identification of ordinal position. J. Exp. Psychol. Anim. Behav. Process. 33, 21–31.
- Rugani, R., Regolin, L., Vallortigara, G., 2010. Imprinted numbers: newborn chicks' sensitivity to number vs. continuous extent of objects they have been reared with. Dev. Sci. 13, 790–797.
- Santiago, H.C., Wright, A.A., 1984. Pigeon memory: same/different concept learning, serial probe recognition acquisition, and probe delay effects on the serialposition function. J. Exp. Psychol. Anim. Behav. Process. 10, 498–512.
- Savage-Rumbaugh, E.S., Rumbaugh, D.M., Boysen, S., 1980. Do apes use language? Am. Sci. 68, 49-61.
- Savage-Rumbaugh, S., Murphy, J., Sevcik, R.A., Brakke, K.E., Williams, S.L., Rumbaugh, D.M., 1993. Language comprehension in ape and child. Monogr. Soc. Res. Child Dev. 233, 1–258.
- Scarf, D., Hayne, H., Colombo, M., 2011. Pigeons on par with primates in numerical competence. Science 334, 1664.
- Siegel, L.S., 1982. The development of quantity concepts: perceptual and linguistic factors. In: Brainerd, C.J. (Ed.), Children's Logical and Mathematical Cognition. Springer-Verlag, Berlin, Heidelberg, New York, pp. 123–155.
- Skinner, B.F., 1938. The Behavior of Organisms. Appleton-Century-Crofts, New York. Starkey, P., Cooper, R.G., 1995. The development of subitizing in young children. Brit. I. Dev. Psychol. 13, 399–420.
- Teubal, E., Guberman, A., 2002. The development of children's counting ability. Megamot 42, 83–102.
- Tinbergen, N., 1951. The Study of Instinct. Oxford University Press, New York.
- Todt, D., 1975. Social learning of vocal patterns and modes of their applications in Grey parrots. Z. Tierpsychol. 39, 178–188.
- Trick, L., Pylyshyn, Z., 1989. Subitizing and the FNST Spatial Index Model. University of Ontario, Ontario, Canada, COGMEM #44.
- Trick, L., Pylyshyn, Z., 1994. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychol. Rev. 101, 80–102.
- Vallortigara, G., 2012. Core knowledge of object, number, and geometry: a comparative and neural approach. Cogn. Neuropsychol., http://dx.doi.org/ 10.1080/02643294.2012.654772.
- von Glasersfeld, E., 1982. Subitizing: the role of figural patterns in the development of numerical concepts. Arch. Psychol. 50, 191–218.
- Watanabe, S., Huber, L., 2006. Animal logics: decision in the absence of human language. Anim. Cogn. 9, 235–245.
- Wellman, H.M., Miller, K.F., 1986. Thinking about nothing: development of concepts of zero. Brit. J. Dev. Psychol. 4, 31–42.
- Wright, A.A., Santiago, H.C., Sands, S.F., 1984a. Monkey memory: same/different concept learning, serial probe acquisition, and probe delay effects. J. Exp. Psychol. Anim. Behav. Process. 10, 513–529.
- Wright, A.A., Santiago, H.C., Urcuioli, P.J., Sands, S.F., 1984b. Monkey and pigeon acquisition of same/different concept using pictorial stimuli. In: Commons, M.L., Herrnstein, R.J., Wagner, A.R. (Eds.), Quantitative Analysis of Behavior, 4. Ballinger, Cambridge, MA, pp. 295–317.
- Wynn, K., 1990. Children's understanding of counting. Cognition 36, 155–193.
- Wynn, K., 1992. Children's acquisition of the number words and the counting system. Cogn. Psychol. 24, 220–251.

Contents lists available at SciVerse ScienceDirect





Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Nothing to it: Precursors to a zero concept in preschoolers

Dustin J. Merritt^a, Elizabeth M. Brannon^{a,b,*}

^a Center for Cognitive Neuroscience, Duke University, United States
^b Department of Psychology and Neuroscience, Duke University, United States

ARTICLE INFO

Article history: Received 29 August 2012 Received in revised form 30 October 2012 Accepted 2 November 2012

Keywords: Zero Empty-set Distance effect Numerical cognition Numerical continuum

ABSTRACT

Do young children understand the numerical value of empty sets prior to developing a concept of symbolic zero? Are empty sets represented as mental magnitudes? In order to investigate these questions, we tested 4-year old children and adults with a numerical ordering task in which the goal was to select two stimuli in ascending numerical order with occasional empty set stimuli. Both children and adults showed distance effects for empty sets. Children who were unable to order the symbol zero (e.g., 0 < 1), but who successfully ordered countable integers (e.g., 2 < 4) nevertheless showed distance effects with empty sets. These results suggest that empty sets are represented on the same numerical continuum as non-empty sets and that children represent empty sets numerically prior to understanding symbolic zero.

© 2012 Elsevier B.V. All rights reserved.

Young children face a difficult cognitive challenge when learning the meaning of number words and the verbal count procedure (e.g., Fuson and Hall, 1983; Wynn, 1990, 1992; Carey, 2009). One of the many reasons why this is such a difficult process for children is that numbers are abstract symbols that are not bound to the physical and perceptual qualities of a stimulus set. For example, three trees and three cars differ in many important perceptual features, but they both share the common abstract feature of "threeness". Other numbers that are not part of the count-list, such as zero, present an even greater challenge. Unlike the integers, zero does not represent the presence of a specific quantity; rather, it represents the absence of a quantity. As a result, the zero concept may present unique developmental and conceptual challenges for children.

The introduction of zero into modern symbolic notational systems occurred long after the incorporation of the count-list numbers. One of the earliest uses of zero was by the Babylonians (approximately 1500 BCE), who used zero as a placeholder to indicate the absence of a particular numerical value (e.g., zero in the number 101 represents the absence of a value in the "tens" column). Later, the Greeks used zero to indicate "absence", but only centuries later, in India, was zero introduced as a number that could be used in mathematical computations (Bialystok and Codd, 2000; Menninger, 1992).

The relatively delayed introduction of zero into modern notational systems is paralleled by the delayed developmental trajectory of mastering the meaning of the word "zero" in young children. Wellman and Miller (1986) demonstrated that children master the cardinal and ordinal properties of the count-list integers before they incorporate "zero". In one experiment, children were presented with an array of four cubes and asked to first count the cubes and then count backwards as one cube at a time was removed from the array. They found that children were much better at verbally identifying the number of cubes when there were more than zero cubes compared to when there were no cubes remaining. Similarly, when asked to make numerical magnitude comparisons, children were much more accurate comparing the numerals 1–5 with each other than comparing the count-list integers with zero.

Children's understanding of the symbol zero develops in a series of stages (Wellman and Miller, 1986). Children first learn to identify the symbol for zero without understanding what the symbol means. Later, children learn that zero represents "none" or "nothing", but, they still fail to recognize that zero is a numerical value that occupies a place on the numerical continuum. For example, when asked, "which is smaller, zero or one?" children will often insist that "one" is the smaller number (Wellman and Miller, 1986). Finally, children learn the relationship between zero and the other numbers on the continuum, and appreciate that zero is smaller than one. Interestingly, confusion surrounding the zero concept is not unique to children; even educated adults have difficulty grasping the status of zero as a number, and how zero operates in mathematical calculations (Wheeler and Feghali, 1983).

The delay in children's understanding of zero raises an important question. Do children incorporate zero into the same numerical

^{*} Corresponding author at: Duke University, Center for Cognitive Neuroscience, B203 LSRC, Box 90999, Durham, NC 27708, United States. Tel.: +1 919 668 6201; fax: +1 919 681 0815

E-mail address: Brannon@duke.edu (E.M. Brannon).

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.001

continuum that is occupied by the count-list integers? Or do they initially develop a fundamentally different representation for zero, and only learn to evaluate it relative to other numbers through a system of rules?

1. Evidence that zero is represented differently than the count-list

Positive integers may be psychologically privileged because they can be represented as magnitudes by a numerical accumulator (Wynn, 1998; Meck and Church, 1983). Wynn argued that preverbal children and animals are unable to represent empty sets given their reliance on an accumulator to represent numerosities. There is no mental magnitude value for zero in such an accumulator.

Consistent with this line of reasoning, Wynn and Chiang (1998) found that 8-month old infants tested in a modified version of the well-known addition and subtraction paradigm developed by Wynn (1992) did not form expectations of empty set outcomes. In a magical disappearance condition, infants watched as an object was placed behind a screen. The experimenter then secretly removed the object, so that it appeared to have disappeared when the screen was lowered. In a magical appearance condition, a screen was raised in front of a single object while the infant watched. The experimenter then removed the object in view of the infant, but secretly replaced the object behind the screen. Thus, when the screen was lowered, the object seemed to have magically appeared. When compared to control conditions, infants looked longer at the magical disappearance, but did not look longer at the magical appearance. Wynn and Chiang argued that the infants did not notice the magical appearance because they were unable to represent the initial empty set, and therefore did not form the expectation that there should be no objects behind the occluder.

Experiments with adults also suggest that zero is represented differently than other natural numbers. Brysbaert (1995) examined numerical processing time for the Arabic numerals 0-99. Participant's eye movements were tracked as they sequentially viewed three numerals presented on a screen. After viewing all three numerals, participants answered whether the middle numeral in the sequence had a magnitude between the first and third value (e.g., 21, 23, 27), or, whether its magnitude was outside the first and third value (e.g., 27, 21, 23). Brysbaert found that the amount of time participants looked at each numeral could be predicted by the numeral's logarithmic value. However, the amount of time participants looked at zero was significantly longer than the amount of time they looked at the numeral "one". Brysbaert concluded that a single system could account for scan patterns to numerals greater than zero, but that zero itself was treated qualitatively differently and was not part of the mental number line.

Other differences between zero and the counting numbers emerge when examining symbolic arithmetic (Butterworth et al., 2001). Wellman and Miller (1986) report that, unlike the counting integers, young children rely on simple algebraic rules when making calculations with zero such as n + 0 = n and n - 0 = n. Wellman and Miller speculate that children's difficulty in understanding zero might contribute to their rule-based approach to mathematical operations. Similarly, other researchers have found that some patients with brain damage show deficits for fact-based arithmetic processing (e.g., $7 \times 3 = 21$) that differ from those of rule-based arithmetic processing involving zero (e.g., $N \times 0 = 0$) (Mccloskey et al., 1991; Semenza et al., 2006). However, it is difficult to draw conclusions about how zero is represented from studying arithmetic operations given that rule-based calculations can be used even without a strong conceptual understanding of zero (Semenza et al., 2006).

1.1. Evidence that zero is represented as part of the count-list

Although the previously mentioned studies suggest that empty sets and the symbol zero may be represented differently than the other numbers, some evidence suggests that zero is represented along a common continuum with positive integers. Bialystok and Codd (2000) placed 5 dolls representing "Sesame Street" characters on a table. In front of each doll were two boxes, one box for lunch, and the other for an afternoon snack. Children, aged 3–7 years old, were asked to distribute 2, 5, 0, 1/2, or 1/4 cookies in each box. For example, the experimenter might say: "Can you give Big Bird 2 cookies for lunch?" Children were than asked to label the boxes with post-it notes so they could later identify the number of cookies in the boxes. Consistent with the findings of Wellman and Miller (1986), three-year-old children had more difficulty distributing zero cookies than whole number cookies. However, children also showed some remarkable similarities in their labeling methods for both zero and whole numbers. Three and four year olds were more likely to use iconic symbols than were older children. Specifically, they drew circles to represent cookies on a single postit in one-to-one correspondence to the number of cookies that were placed into the box. For example, a child who was asked to place two cookies in a box labeled the quantity by drawing two circles on the post-it note. For boxes that contained zero cookies, young children left the post-it blank to represent the absence of cookies in the box

Other evidence suggests that empty sets can be represented by non-human animals as analog magnitudes. Given that previous research has shown that humans and animals share an approximate number system (ANS) for processing numerical values, recent findings in animals may be particularly relevant for exploring nonverbal numerical processing in young children (Brannon and Terrace, 2000; Cantlon and Brannon, 2006; Feigenson et al., 2004; Gelman and Gallistel, 2004; Nieder and Miller, 2004). When comparing the relative magnitude of numerical arrays, humans and animals show similar distance and magnitude effects. Generally, as the disparity (distance) increases between two numerical sets, discrimination becomes easier. When the distance between two numerical sets is held constant, but the magnitude of those sets increases (e.g., 2 and 4 vs. 8 and 10), discrimination becomes more difficult. Therefore, if empty sets occupy a place on the numerical continuum, then when comparing empty sets with other values, we would expect lower accuracy when the distance from the empty set is small and greater accuracy when the distance from the empty set is large.

Biro and Matsuzawa (2001) trained a chimpanzee (named Ai) to match arrays of dots to corresponding Arabic numerals. The numeral zero was matched to a blank square that did not contain any dots. Ai was then tested with an ordinal task in which she was required to select the Arabic numbers in order from smallest to largest. Ai was unable to spontaneously transfer the symbol zero from the matching task to the ordinal task initially; although not surprisingly, she eventually learned to correctly order the symbol zero relative to the other symbols. In another set of studies, Pepperberg and colleagues found that an African Gray Parrot (named Alex) spontaneously used the word "none" to identify absence in some numerical contexts, but not in others (Pepperberg and Gordon, 2005; Pepperberg, 2006).

Although results with Ai the chimpanzee and Alex the parrot suggest that animals may possess some important features of a zero-like concept, both animals showed limitations in their ability to transfer a zero symbol to a novel context. Would animals fare better if the symbolic requirement were removed? To address this question and to assess whether animals have precursors to a zero concept when not required to learn arbitrary stimuli. Merritt et al. (2009) tested rhesus monkeys, who were already

proficient at numerical matching and ordering tasks, on their ability to respond to empty sets. The matching task required monkeys to select a target that numerically matched a previously shown sample array. The numerical ordering task required monkeys to respond to two numerical arrays in ascending (red background) or descending (blue background) order. In both tasks, standard trials contained numerical arrays with values between 1 and 12. Correct answers produced positive (juice) feedback and incorrect answers produced negative (timeout) feedback. On probe trials, monkeys were tested on their ability to match or order empty set stimuli that consisted of a blank square that varied in color and size but contained no elements. In order to prevent the monkeys from learning the correct choice during probe testing, the probe-trial choices were not differentially reinforced. Results indicated that the monkeys were able to spontaneously match and order empty sets at accuracy levels comparable to those of the other numerosities. Further, in both tasks the monkeys showed distance effects, with accuracy increasing as the distance between the empty set and the other numerosity increased. This pattern of results suggests that the monkeys were not treating the empty set as a qualitatively different non-numerical stimulus, but rather, they viewed the empty set a numerical value that could be directly compared to other numerical values.

Here, we tested four-year old children and adults using procedures similar to those of Merritt et al. (2009). Four-year old children were chosen because this is the age where children start to resolve their confusion about zero, and begin to understand both its cardinal and ordinal properties (Wellman and Miller, 1986). Our study was designed to answer two main questions. First, are mental magnitudes generated for empty sets? If so, then children and adults should show distance effects indicating representational continuity with the other numerosities. Alternatively children may view empty sets as a qualitatively different nonnumerical stimulus. Second, our study investigated the relationship between a child's ability to order empty sets and their developing understanding of the meaning of the symbol zero. If the ability to order and/or match empty sets serves as a foundation for learning the meaning of the symbol zero, we may see this capacity emerging before reliable usage of the symbol zero in young children.

2. Experiment 1

2.1. Methods

2.1.1. Subjects

Participants were 21 four-year old children (mean age = 4.5, SD = .32). One additional child was excluded from the analyses because he/she avoided selecting the empty set on 96% of all probe trials.

2.1.2. Apparatus

Participants were tested in a small room while seated in front of a 17-in. computer screen affixed with a *MagicTouch* touch sensitive screen. A custom-built program written in RealBasic presented the stimuli and registered the responses.

2.1.3. Procedure

The task required children to select the numerically smaller of two numerosities¹ (see Fig. 1). At the beginning of each trial, a start stimulus (a picture of a white rabbit) appeared in the lower right corner of the screen. Pressing the start stimulus resulted in



Fig. 1. Example screen shots for training and probe trials. The task was to select the smaller numerosity first and the larger numerosity second.

the presentation of two numerical stimuli in random locations on the screen. If the child correctly selected the smaller numerosity, a black border appeared around the stimulus, and the stimulus remained on the screen until the second stimulus was selected. Once the child selected the second stimulus correctly, she received computer generated visual and auditory feedback (a 1-s audio clip and a picture of a sun). If the larger numerosity was selected first, the child received negative visual and auditory feedback (a 1-s "Try Again" audio clip and a 3-s black screen) and the trial ended. Stimuli were yellow squares within which circular elements were randomly placed. Elements varied in size, shape, and color so that none of these dimensions could be used as ordering cues.

Training and testing. The stimuli were trial-unique exemplars of 0, 1, 2, 4, and 8. For any trial that contained an empty set, selecting the stimuli in either ascending or descending order produced positive visual and auditory feedback. This was done so that no information about the empty set was conveyed during the experiment. All possible pairs were presented in a pseudo-random order, with the constraint that each pair was presented an equal number of times throughout the session. The children were given a total of 60 trials.

Controls. In order to eliminate the possibility that children were using background surface area as a cue on empty set trials density was controlled by varying the background size of the two stimuli. The larger numerosity had the larger background size (12.73 cm \times 10.6 cm) on half of the trials and the smaller background size (6.36 \times 5.32 cm) on the other half of the trials.

Instructions and demonstration. Prior to each session, children watched as the experimenter demonstrated the task. The children were instructed to touch the picture with the smaller number of objects first, and to touch the picture with the larger number of objects second. Children were further instructed that they could take as much time as they needed to press the start stimulus, but afterward, they should respond as quickly as possible without counting. On two occasions a child attempted to count aloud, she/he was interrupted and reminded not to count. There were a total of 8 practice trials. Following the demonstration by the experimenter, the child was allowed to complete the remainder of the practice trials. On three occasions when the child failed to get the last two trials correct, the child was given four additional practice trials. The practice trials did not contain the empty set.

¹ In research with adults and monkeys performance has been shown to be equivalent for choosing the larger or smaller array (e.g., Cantlon and Brannon, 2005).

2.2. Symbolic number assays

2.2.1. Children were tested with four post-test numerical assays to assess their knowledge of symbolic numbers

Give-a-number and how many tasks. The "Give a number" and "How many?" tasks were based on Wynn (1992). During the "how many" task, the experimenter placed a set of 1–6 plastic dolphins in a line in front of the child. The child was presented with a different set size on each trial until all 6 sets had been presented. During each trial, the child was asked "How many dolphins are there? Can you count them out loud?" In the "give a number" task, the standard titration procedure was used. Twelve plastic dolphins were placed in a pile, and children were first asked "Can you give me one dolphin?" The number of dolphins requested was increased by one each time the child gave the requested number correctly and was decreased by one each time the child erred. The task was completed when the child successfully gave N dolphins twice and was unsuccessful on two requests for N+1. For both tasks, the experimenter encouraged the children to count and to check their answers by asking "Can you count them for me to make sure?"

Smallest number query. Children were asked, "What is the smallest number in the world?" If the child identified a positive integer, the child was asked "Is there a number smaller than that?" This process was repeated until the child failed to provide a number that was smaller than their previous answer.

Symbolic ordering task. Children were shown two $3 \text{ in.} \times 3 \text{ in.}$ index cards with Arabic numbers written on them. The child was asked" which one is the smaller number?" The values shown were 1 vs. 6, 2 vs. 4, 0 vs. 6, and 0 vs. 1. The left–right positions of the cards were randomized, and pair comparisons were chosen randomly by shuffling the pairs before each session.

2.3. Scoring the symbolic number assays

How many. Scoring for the "how many" and "give a number tasks" followed the procedure used by Cantlon et al. (2007). In the "how many" task, children received 3 points if they were able to correctly count all 6 dolphins and correctly indicate the cardinal number. They received 2 points if they counted a set of six incorrectly, but were able to correct the mistake. They also received 2 points if they counted correctly, but failed to accurately report the total number of dolphins when asked "how many?" They received 1 point if they were able to count at least two dolphins. Zero points were given if the child was unable to count and label correctly for sets of 2 dolphins.

Give a number. For the "give a number task", children received 3 points if they were able to give the experimenter exactly 6 dolphins. They received 2 points if they gave the experimenter an incorrect number of dolphins when asked for 6, but were able to correct their mistake. Children were given 1 point if they were able to accurately hand the experimenter at least two dolphins, and zero points if they were unable to produce at least two dolphins correctly.

Smallest number and symbolic ordering. The "smallest number" and "symbolic ordering" tasks were scored in a binary fashion, as either pass or fail. In order to pass the smallest number task, the child had to answer that zero was the smallest possible number. If the child answered with a number greater than zero, then the child was scored as having failed the task. In the symbolic ordering tasks, comparisons involving zero were scored separately from non-zero comparisons. For zero comparisons, children who identified zero as the smallest number for *both* comparisons were scored as passing (0 vs. 1, and 0 vs. 6). In contrast, children who chose the larger number in either comparison were scored as failing. Similarly, children who identified the smallest number for both non-zero comparisons



Fig. 2. A comparison of ordering performance on empty set and non-empty set trials. Points in the shaded area indicate greater accuracy on empty sets than on non-empty sets. Points in the unshaded area indicate greater accuracy on non-empty set than on empty set trials.

(1 vs. 6, and 2 vs. 4) were scored as passing, whereas children who did not answer both correctly were scored as failing.

3. Results

3.1. Symbolic performance

On the "how many" task, 75% of all children scored the maximum score of 3, 25% scored a 2, and no children scored a 0 or 1. On the "give a number" task, 75% of children scored the maximum score of 3, 15% scored a 2, and 10% scored a 1. Further, 75% of children accurately ordered both pairs of positive numbers (binomial, p < .05), and 45% accurately ordered both pairs containing zero (binomial, p < .05). A Wilcoxon signed ranks test revealed that children had more difficulty ordering comparisons involving zero than comparisons of positive integers (Z = -2.12, p < .05). When tested on the "smallest number" task, 57% of children identified zero as the smallest number; the remaining children identified numbers larger than zero.²

3.2. Non-symbolic performance

Overall, children ordered (non-empty) numerical sets at above chance accuracy [M=73.9%; t(19)=5.9, p <.05]. However, children were considerably more accurate ordering non-empty numerical sets compared to ordering pairs that contained an empty set (Fig. 2); paired samples *t*-test [t(19)=3.12, p <.05]. Empty set accuracy did not exceed chance for pairs 0,1, [t(19)=0, p=1.0] and 0,2 [t(19)=1.25, p=0.12] but did exceed chance for pairs 0,4 [t(19)=1.74, p <.05] and 0,8 [t(19)=2.89, p <.05]. Due to large variability and inconsistencies in RT performance, we did not analyze RT for distance effects (e.g., M=2.09 s, SD=1.65 s).

Because empty sets do not lend themselves to ratio comparisons, we compared distance effects for empty set probe trials with standard trials that contained the numerosity one (see also Merritt et al., 2009). If children treat empty sets as numerical values that can be compared with other numerical values, then we would expect that distance effects with empty sets should be similar to those observed with sets of one. As demonstrated in Fig. 3, children showed distance effects for both empty sets F(3,57)=3.88, p < .05, and for numerical comparisons involving one item, F(2,38)=7.22, p < .05. The empty set distance effects shown here are qualitatively similar to those found by Merritt et al. (2009) in rhesus monkeys.

² One child failed to provide an answer on the "smallest number" task.



Fig. 3. Accuracy for empty sets and sets containing one item as a function of distance. Also included for comparison purposes, is the empty set accuracy across distances of 1, 2, 4, and 8 previously obtained from rhesus monkeys. The monkey data are from Merritt et al. (2009).

If empty sets are represented as mental magnitudes, then children who are more proficient at forming representations of positive numerosities should also be more proficient at forming representations of empty sets. In order to assess whether proficiency ordering sets of one translated into proficiency ordering empty sets, we divided children based on a median split according to their accuracy on standard trials that included the numerosity one. We then examined accuracy for empty sets as a function of distance in each of these groups.

Children were divided into two groups based on their performance on numerosity comparisons that did not include empty sets. The children who were in the bottom half of this group were above chance on numerosity comparisons with sets of one, t(9)=2.51, p < .05, but were at chance on comparisons with empty sets, t(9)=-0.04, p=0.49. Further these children did not show distance effects with empty sets (F(3,27)=1.05, p=0.39) or for sets of one. While a repeated-measures ANOVA on distance for comparisons with the numerosity one was significant, F(2,18)=8.67, p < .05; trend tests revealed a quadratic function, F(1,9)=23.65, p < .05, rather than a linear function F(1,9)=0.34, p=0.58 reflecting that accuracy was highest for the middle distance rather than the largest distance.

For top half performers, accuracy was above chance for both empty sets t(9) = 2.17, p < .05, and sets of one, t(9) = 13.28, p < .05, and, they showed reliable distance effects for both empty sets, F(3,27) = 3.59, p < .05, and sets of one F(2,18) = 6.77, p < .05.

3.3. Relationship between performance on empty sets and symbolic number knowledge

Fifteen of the twenty children scored at ceiling on the "how many" and "give a number" tasks (not the exact same 15 children). For each task, accuracy on empty sets (e.g., choosing empty sets before a comparison array with 1–9 elements) was better for the 15 children who scored at ceiling compared to those children who scored lower (60% vs 52% and 61% vs 54% for the "give a number" and "how many" tasks, respectively). Given the small and uneven sample sizes, neither of these comparisons were significantly different.

The final assay required knowledge of Arabic numerals rather than number words. Children who failed to order both pairs with the symbol zero were less accurate ordering empty sets (M = 50%)



Fig. 4. Empty set accuracy for children who failed to correctly order the zero symbolic number pairs as a function of whether they correctly or incorrectly ordered symbolic number pairs containing positive symbolic (non-zero) numbers.

than children who successfully ordered both pairs with the symbol zero (M=71%; t(18)= -1.91, p<.05). Similarly, children who identified zero as the smallest number were also significantly more accurate ordering empty sets (M=74%) compared to children who did not identify zero as the smallest number (M=43%, t(17)=2.9, p<.05). Thus, performance on symbolic zero tasks was generally predictive of performance on ordering empty sets.

We examined performance on empty sets for the children who correctly ordered both pairs of positive numbers, but failed to order the two pairs with symbolic zero. Overall performance for these children did not exceed chance levels on empty sets [M=57%]; t(6) = 0.75, p = 0.24]. Curiously however, as shown in Fig. 4, these children did show a distance effect, with accuracy increasing as distance between the empty set and the other numerosity increased. F(3,18) = 3.77, p < .05. This suggests that the children may actually possess a rudimentary understanding of how empty sets relate to other numerosities. To explore this possibility we compared empty set performance for children who were unable to successfully order the symbol zero, but were able to order the positive numbers (n = 7), with children who failed both (n = 5). Unlike children who were able to order positive numbers, children who failed both tasks did not show a distance effect for empty sets [F(3,9)=0.16, p=0.92]. Further, their performance was significantly lower than children who correctly ordered the positive numbers [t(10) = -1.9, p < 05], and as shown in Fig. 4, their overall accuracy was significantly below chance, thereby demonstrating a bias against selecting empty sets [M=36%, t(4)=-2.5, p<.05].

Overall, these results suggest that children may have a burgeoning understanding of the ordinal relationship between empty sets and other numerosities before they understand how the symbol zero relates to other symbolic numbers. They also suggest that comprehension symbolic zero's numerical meaning is unlikely to be critical in and of itself for children to appreciate empty sets as magnitudes on a mental number line. We return to the issue of the relationship between symbolic number knowledge and empty set performance in the discussion.

4. Experiment 2

When human adults are tested in numerical tasks that avoid verbal counting, their accuracy and RT is ratio dependent and appears to tap an approximate number system (ANS) shared with a variety of nonhuman animals (Cantlon and Brannon, 2006; see also Cordes et al., 2001; Platt and Johnson, 1971; Whalen et al., 1999). Given that zero is not a countable number, it may be that it is appreciated



Fig. 5. Reaction time as a function of distance for adult participants.

as categorically different from the integers. Certainly, as explained earlier, some evidence suggests that zero is not represented by the same way as other numbers (Brysbaert, 1995). Thus Experiment 2 investigates whether human adults represent empty sets as analog numerical magnitude values, and thus show distance effects or alternatively whether they treat empty sets as categorically different from other numerosities.

4.1. Subjects

The participants were 10 undergraduate students from Duke University who participated in exchange for payment.

4.2. Procedure and apparatus

The procedure and apparatus were identical to that of the ordinal task in Experiment 1 with a few exceptions. First, rather than using a touch screen, participants responded by clicking each stimulus with a mouse. Second, participants were given 540 trials with all numerosities 0–9 within a single session. Third, correct choices were rewarded with a white screen that displayed the word "Correct!" in black letters. Incorrect choices produced a black screen with the word "Incorrect!" in white letters. Like the children, adults were rewarded for both correct and incorrect responses made during the empty set trials. No symbolic numerical assays were given to the adult participants. There was no time limit, but participants were asked to select the numerically smaller array as quickly and accurately as they could without counting.

5. Results

As with the previous experiments, we compared distance effects for empty set probe trials with standard trials that included the numerosity one. Overall accuracy was extremely high for empty set probe trials (M=98.5%) as well as trials containing the numerosity one (M=98.8%). There were no difference in accuracy between empty set probe trials and standard trials that included the numerosity one [t(9)=-0.54, p=0.60]. No further analyses were done on accuracy given that it was at ceiling levels.

Fig. 5 shows that adults exhibited a distance effect in RT for both empty sets and standard trials with the numerosity one. A repeated measures one-way ANOVA revealed that RT on correct trials decreased as distance increased for comparisons involving the empty set [F(8,72)=2.78, p<.05] and those involving a set of one [F(7,63)=3.93, p<.05]. Further, a two-way ANOVA with factors of distance (1–8) and comparison type (empty set vs. one) showed that there was no difference in the slope between the two comparison types [F(7,63) = 0.92, p = 0.50].

5.1. General discussion

Our results suggest that children as young as four years of age represent empty sets along the same numerical continuum as other numerosities and thus represent them as analog magnitudes. At a group level children showed distance effects for comparisons with empty sets that were similar to reaction-time patterns in human adults (Experiment 2) and accuracy in monkeys (Merritt et al., 2009). However, there was variability in children's ability suggesting that the representation of empty sets may be in flux at 4 years of age.

When children were divided into two groups based on their performance on numerosity comparisons that did not include the empty sets, the higher performing group showed high accuracy levels and distance effects with empty sets. In contrast, the lower performing group was at chance on empty set comparisons and did not show a distance effect. Taken together, these findings suggest that children are incorporating empty sets into their mental number line in this developmental window.

A second question we investigated was whether young children understand the ordinal relationship between empty sets and the other numerosities before they understand the ordinal properties of the symbol for zero. We found that children who successfully ordered all four symbol pairs (two with zero and two without), were above chance on their ability to order empty sets and showed a distance effect for empty sets. Children who successfully ordered both of the positive integers but failed to order both of the pairs with the symbol zero were at chance on empty sets but nevertheless showed a distance effect for empty sets. Children, with the most limited knowledge of numerical symbols, who failed to correctly order one or more within each of the two trial types performed at chance on empty sets and did not show a distance effect for empty sets.

One possible explanation of these results is that two different factors are contributing to children's developing empty set performance. The first is an initial bias to avoid empty sets and the second is increasing understanding that empty sets represent a value less than one on the mental number line. Children may have an initial bias against selecting the empty set. This bias may emerge from children's experience with a verbal count list, which starts with "one." This may create top-down interference that disrupts their ability to order empty sets. Parents rarely draw a child's attention to the absence of countable objects nor do they frequently start counting with the word "zero". This enculturation may lead children to actively avoid empty sets. This avoidance may decrease as empty sets become fully incorporated into the child's mental number line, which may be occurring gradually as a separate process.

The protracted emergence of an appreciation of empty sets as numerical entities is consistent with Wynn and Chiang's (1998) finding that infants do not represent empty sets. However, the fact that empty sets are indeed treated as analog magnitudes by 4 years of age and into adulthood (in the current context) suggests that the nonverbal system for representing number as analog magnitudes is in fact capable of a zero setting. Acquisition of the symbol zero appears to occur in concert with, but is slightly delayed relative to their appreciation of the numerical value of empty sets. This raises the interesting question of how symbolic knowledge contributes to the non-symbolic appreciation of empty sets. What kinds of experience are most critical? Our study cannot speak to the question of whether or how developing ordinal symbolic knowledge informs a child's understanding of nonsymbolic empty sets. But one possibility the research suggests is that it may be necessary for children who do not yet understand the ordinal properties of symbolic zero to learn the ordinal relationship between the positive (non-zero) numbers prior to incorporating empty sets into the mental number line.

6. Conclusions

Prior work from our research group has demonstrated that monkeys represent empty sets as a value along the numerical continuum (Merritt et al., 2009). This finding inspired us to ask whether adult humans and young children also represent empty sets in an analog fashion. We found unequivocal evidence that adults represent empty sets as mental magnitudes. However, our findings with 4-year-old children were more variable. In both tasks, only children who successfully ordered countable numerosities showed distance effects for empty sets. Thus consistent with the claim that infants may be unable to represent empty sets numerically we find that the capacity for appreciating empty sets as values along the numerical continuum has a protracted development (Wynn and Chiang, 1998). Further research, with younger children will be necessary to pinpoint when representing empty sets on the numerical continuum emerges over development and whether mapping of numerical symbols to numerosities plays an important role in empty set representations.

Acknowledgements

This paper is contributed as a tribute to Dr. Tony Wright and his research program. While neither of the authors had the opportunity to work with Tony as a student, our research programs have been greatly influenced by his contributions to the field of comparative cognition. This work was supported by an NSF Career award Grant No. 0448250, a James McDonnell scholar award to EMB, and by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award No. F32MH082623 to Dustin J. Merritt. We also thank Maggie Vogel for assisting with this research and the members of the Brannon lab for their helpful discussion of these data.

References

- Bialystok, E., Codd, J., 2000. Representing quantity beyond whole numbers: some, none, and part. Can. J. Exp. Psychol. 54, 117–128.
- Biro, D., Matsuzawa, T., 2001. Use of numerical symbols by the chimpanzee (*Pan troglodytes*): cardinals, ordinals, and the introduction of zero. Anim. Cogn. 4, 193–199.

- Brannon, E.M., Terrace, H.S., 2000. Representation of the numerosities 1–9 by rhesus macaques (Macaca mulatta). J. Exp. Psychol. Anim. Behav. Process. 26 (1), 31–49.
- Brysbaert, M., 1995. Arabic number reading on the nature of the numerical scale and the origin of phonological recoding. J. Exp. Psychol. Gen. 124 (4), 434–452. Butterworth, B., Zorzi, M., Girelli, L., Jonckheere, A.R., 2001. Storage and retrieval of
- addition facts: the role of number comparison. Quart. J. Exp. Psychol. A: Hum. Exp. Psychol. 54 (4), 1005–1029.
- Cantlon, J., Brannon, E.M., 2005. Semantic congruity facilitates number judgments in monkeys. Proc. Natl. Acad. Sci. 102 (45), 16507–16511.
- Cantlon, J., Fink, R., Safford, K., Brannon, E.M., 2007. Heterogeneity impairs numerical matching but not numerical ordering in preschool children. Dev. Sci. 10 (4), 431–440.
- Cantlon, J.F., Brannon, E.M., 2006. Shared system for ordering small and large numbers in monkeys and humans. Psychol. Sci. 17 (5), 402–407.
- Carey, S., 2009. The Origin of Concepts. Oxford University Press, New York.
- Cordes, S., Gelman, R., Gallistel, C., Whalen, J., 2001. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. Psychon. Bull. Rev. 8 (4), 698–707.
- Feigenson, L., Dehaene, S., Spelke, E., 2004. Core systems of number. Trends Cogn. Sci. 8 (7), 307–314.
- Fuson, K.C., Hall, J.W., 1983. The acquisition of early number word meanings: a conceptual analysis and review. In: Ginsburg, H.P. (Ed.), The Development of Mathematical Thinking. New York, Academic Press.
- Gelman, R., Gallistel, C.R., 2004. Language and the origin of numerical concepts. Science 306 (5695), 441–443.
- Mccloskey, M., Aliminosa, D., Sokol, S.M., 1991. Facts, rules, and procedures in normal calculation – evidence from multiple single-patient studies of impaired arithmetic fact retrieval. Brain Cogn. 17 (2), 154–203.
- Meck, W.H., Church, R.M., 1983. A mode control model of counting and timing processes. J. Exp. Psychol. Anim. Behav. Process. 9 (3), 320–334.
- Menninger, K., (Z. Ziffer, Trans.) 1992. (Number words and number systems). Dover, Mineola.
- Merritt, D.J., Rugani, R., Brannon, E.M., 2009. Empty sets as part of the numerical continuum: conceptual precursors to the zero concept in rhesus monkeys. J. Exp. Psychol. Gen. 138 (2), 258–269.
- Nieder, A., Miller, E.K., 2004. Analog numerical representations in rhesus monkeys: evidence for parallel processing. J. Cogn. Neurosci. 16 (5), 889–901.
- Pepperberg, I.M., 2006. Grey parrot (*Psittacus erithacus*) numerical abilities: addition and further experiments on a zero-like concept. J. Comp. Psychol. 120 (1), 1–11. Pepperberg, I.M., Gordon, J.D., 2005. Number comprehension by a grey parrot (*Psitta*-
- *cus erithacus*), including a zero-like concept. J. Comp. Psychol. 119 (2), 197–209. Platt. I.R. Johnson, D.M. 1971. Localization of position within a homogeneous behav-
- ior chain: Effects of error contingencies. Learn. Motiv. 2, 386–414. Semenza, C., Grana, A., Girelli, L., 2006. On knowing about nothing: the processing of
- zero in single- and multi-digit multiplication. Aphasiology 20 (9), 1105–1111. Wellman, H.M., Miller, K.F., 1986. Thinking about nothing: developmental concepts
- of zero. Brit. J. Dev. Psychol. 4, 31–42. Whalen, J., Gallistel, C.R., Gelman, R., 1999. Nonverbal counting in humans: the
- psychophysics of number representation. Psychol. Sci. 10 (2), 130–137.
- Wheeler, M., Feghali, I., 1983. Much ado about nothing: preservice elementary school teachers' concept of zero. J. Res. Math. Educ. 14 (3), 147–155.
- Wynn, K., 1990. Childrens understanding of counting. Cognition 36 (2), 155-193.
- Wynn, K., 1992. Childrens acquisition of the number words and the counting system. Cogn. Psychol. 24 (2), 220–251.
- Wynn, K., 1998. Psychological foundations of number: numerical competence in human infants. Trends Cogn. Sci. 2 (8), 296–303.
- Wynn, K., Chiang, W.C., 1998. Limits to infants' knowledge of objects: the case of magical appearance. Psychol. Sci. 9 (6), 448–455.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Categorization of birds, mammals, and chimeras by pigeons

Robert G. Cook^{a,*}, Anthony A. Wright^b, Eric E. Drachman^a

^a Tufts University, United States

^b University of Texas Health Science Center, Houston, United States

ARTICLE INFO

Article history: Received 27 June 2012 Received in revised form 31 October 2012 Accepted 4 November 2012

Keywords: Pigeons Categorization Chimeras Vision Feature learning

ABSTRACT

Identifying critical features that control categorization of complex polymorphous pictures by animals remains a challenging and important problem. Toward this goal, experiments were conducted to isolate the properties controlling the categorization of two pictorial categories by pigeons. Pigeons were trained in a go/no-go task to categorize black and white line drawings of birds and mammals. They were then tested with a variety of familiar and novel exemplars of these categories to examine the features controlling this categorization. These tests suggested the pigeons were segregating and using the principal axis of orientation of the animal figures as the primary means of discriminating each category, although other categorical and item-specific cues were likely involved. This perceptual/cognitive reduction of the categorical stimulus space to a few visual features or dimensions is likely a characteristic of this species' processing of complex pictorial discrimination problems and is a critical property for theoretical accounts of this behavior.

© 2012 Published by Elsevier B.V.

It is well established that a number of animals can learn to discriminate and categorize a wide variety of ill-defined, open-ended, natural categories. Pigeons, for example, have learned to discriminate "natural" polymorphous noun categories such as flowers, cars, trees, chairs, cats, dogs, and people (Aust and Huber, 2001; Ghosh et al., 2004; Herrnstein, 1979; Herrnstein and Loveland, 1964; Wasserman et al., 1988). Besides supporting rapid learning, these types of categorical discriminations have been established to support transfer to novel exemplars similar to human conceptual behavior. Because of this similarity and its implications for the evolution of cognition, visual discriminations of this type have generated considerable interest since their inception.

One important issue in the analysis of visual categorization centers on what properties control discrimination and transfer performance. A shortcoming in many categorization experiments has been the scarcity of information about the nature of the cues regulating such discriminations. Without knowing what cues or features are being used by the animals, however, it is difficult to make inferences about the representation of these categories, their underlying computational mechanisms, or their similarity to human conceptual behavior (Cerella, 1986; Cook, 1993; Huber, 2001; Lea et al., 2006a). With these issues in mind, this paper describes

E-mail address: Robert.Cook@tufts.edu (R.G. Cook). *URL*: http://www.pigeon.psy.tufts.edu (R.G. Cook).

0376-6357/\$ - see front matter © 2012 Published by Elsevier B.V. http://dx.doi.org/10.1016/j.beproc.2012.11.006 experiments focused on identifying the visual properties controlling the discrimination by pigeons of two representative pictorial noun categories – birds and mammals.

Research on natural categories has relied on photographs as the primary medium for testing such discrimination. A major limitation with this type of complex stimulus is that it is not easily manipulated. While this photographic complexity may be a key element in the formation of such discriminations, they correspondingly make it difficult to isolate the controlling cues. The availability of modern software for manipulating such images has resulted, however, in some progress. The most in-depth analysis of this type has been the series of experiments conducted by Aust and Huber (2001, 2002, 2003) examining the controlling properties involved with categorizing pictures of people from non-people by pigeons. Using a variety of different transformations (e.g., image scrambling and inversion, part deletion, gray scale) the results of these tests highlight the complexity of this analytic goal. Initial tests involving the scrambling of the entire image suggested that local cues associated with the people and image color were particularly critical (Aust and Huber, 2001). Subsequent research suggested that some portions of the human body (heads, hands) were more important than others (Aust and Huber, 2002) and that the spatial configuration of these parts may be at least partially encoded (Aust and Huber, 2003). The importance of the head has also been confirmed by the pecking and tracking of this part in a people-present/ people-absent discrimination (Dittrich et al., 2010). Finally, their results suggested that both item-specific information about the individual exemplars and category-specific information about the class of items were both being encoded by the pigeons as

^{*} Corresponding author at: Department of Psychology, 490 Boston Ave., Tufts University, Medford, MA 02155, United States. Tel.: +1 617 627 2546; fax: +1 617 381 3189.

determined by their different responses to tests with familiar and novel exemplars. Using a similar approach to examining the visual categorization of people, flowers, chairs and cars by pigeons, Lazareva et al. (2006b) found that different visual attributes controlled them. From tests involving stimulus inversion, blurring and scrambling, their results suggested that the categorization of flowers and people were controlled by the overall contour of the images, while cars and chairs were determined by local features.

The analysis of visual categorization by pigeons in this article has its origins in the research of Cook et al. (1990) using stimuli consisting of black and white line drawings of birds and mammals as the discriminative classes. These stimuli were drawn from edited collections (Harter, 1979; Iyari, 1979) of wood cuts and drawings from 19th century scientific journals and popular magazines. These images have the complex characteristics of photographed natural categories, but have several advantages as a medium. First, because of their original scientific and educational purposes, they capture the animals with considerable detail including, key visual features, characteristic or canonical poses and postures, and often include surrounding habitat for each animal. Combined with the considerable irrelevant variation produced by differences in perspective, subjective distance, and the number of animals depicted, these stimuli seemed well suited for the study of categorization. Second, because each image is a simple collection of individual pen strokes, each line can be independently altered allowing for easy manipulation of their features. Third, their black-and-white nature excludes color information. This is valuable because color often overshadows the processing of other features and dimensions of complex stimuli by pigeons. Thus, these stimuli provided an excellent mixture of the featural richness and variation of photographs, with the capacity for easy manipulation.

Cook et al. (1990) established that these line drawings were effective at creating categorical behavior. They found that these line drawings were easy to discriminate, produced robust transfer to novel exemplars of each category, and that speed of learning and degree of transfer varied with the number of training exemplars. Importantly, they also found that the degree of transfer appeared to be sensitive to the similarity of the items within each category as judged from human prototypicality ratings.

The goal of the present research was to identify the controlling features involved in the discrimination of these bird and mammal categories. We used a partitioning strategy to search the possible feature space involving a series of different image manipulations. These manipulations were tested as a pair of tests. The first test involved the manipulation of familiar exemplars while the second tested novel exemplars. This allowed us to assess both item-specific and category-specific information in the pigeons' reaction to the altered stimuli. The pigeons were trained and tested in a go/nogo discrimination task in which they had to discriminate between the categories by pecking at pictures of birds to be reinforced with food, while inhibiting pecking to pictures of mammals that were presented in extinction. Using this established discrimination, we then conducted a series of tests manipulating different aspects of the stimuli. The background, logic and rationale for these tests are described in the next section.

1. Stimulus analytic tests: background, rationale and logic

The purpose of Test 1 was to examine the degree to which the figure of an animal and/or the contextual natural backgrounds/habitats controlled the discrimination. This was important to determine because it has become established that pigeons can memorize the visual content of a very large numbers of pictorial items (Cook et al., 2005; Fagot and Cook, 2006; Vaughan and Greene, 1984). Further, earlier studies had suggested that small differences in the background of photographic images could also be detected and used by the birds (Greene, 1983). Cook et al. (1990) eliminated backgrounds from some of the training stimuli and showed that pigeons had little trouble continuing to discriminate these animal figures without the background, a finding consistent with the hypothesis that the pictured animals were of primary importance to the categorical discrimination. In those experiments, however, the pigeons were not tested with stimuli in which the figures were removed to evaluate how the background itself contributed to the discrimination. In the test conducted here, we removed the background from a larger set of familiar images, and included conditions where the animal in the drawing was removed, leaving only the background. This allowed us to determine whether or not the redundant contextual information contributed to the discrimination.

As detailed below, the results of the first test will show that the animal figure was indeed most important, so we next divided the animal figures into parts, examining the independent contributions of the head, body and legs. Tests 2A and 2B involved using chimera animals involving mixtures of these parts within and across the categories. By mixing and crossing together different portions of each category into a single "chimera" test animal, it was possible to judge which portions of the animal figures were making the greatest contribution to the pigeons' discrimination. Cook et al. (1990) had pilot-tested a few limited examples of such chimera stimuli. The results from three of the four exemplars tested suggested that the body of the animal, rather than features associated with the head, were most important. In the present study, we employ the same strategy but tested greater numbers of chimeras, constructed from a greater variety of animals, to better test and strengthen the conclusions from that earlier study. One set of tests involved exchanging the head and body of the animals from the two categories. The second set of tests involved manipulating the type and number of legs across the two categories.

The next two tests were designed to evaluate the relative contributions and roles of the global organization and local features of the animal figures. In Test 3A the animal figures were divided into three parts involving the head, trunk, and rear sections of the animals. To manipulate global information, conditions were tested in which these different parts were separated from each other by a spatial gap or simultaneously scrambled or inverted from their normal order of appearance. If the order and continuous nature of these different parts were critical, then these alterations to the global organization should be disruptive to the pigeons' performance.

In Test 3B the interior texture was replaced by a solid area of single brightness. This removed local information primarily leaving global shape as the basis for any discrimination. These test stimuli were presented over a range of brightness values, from complete silhouettes, through intermediate brightness values, to exclusively outlined contours. If the global form was exclusively controlling the discrimination, then the pigeons should have little difficulty with these altered forms. On the other hand, if local details in the interior of the animals were also a part of their representation of the categories, then this manipulation should disrupt performance.

Finally, Test 4 examined how the orientation of animal figures influenced the discrimination. Cook et al. (1990) had found that the pigeons were insensitive to either reflections or 180° rotations of the categories, suggesting that orientation was not particularly important. However, both tests had retained the primarily diagonal orientation of the birds and the basic horizontal orientation of the mammals. In Test 4, we included a more extensive and diagnostic set of figural orientations to reexamine the contribution of this global factor to the discrimination of both familiar and novel members of each category.

For purposes of economy, the general methods outline the shared elements of the procedures for the different tests. This is followed by four different sections that include the specific details of each test, along with its results and a brief interpretation. Overall interpretations of the results are then considered in Section 4.

2. General methods

2.1. Animals

Five male White Carneaux pigeons, *Columba livia*, were tested. They were maintained at 80–85% of their free feeding weights with free access to grit and water. The pigeons had previously learned this bird/mammal discrimination 11 months earlier. They were refamiliarized with the discrimination using a combination of old and new exemplars of each category for approximately 1 month before testing commenced.

2.2. Apparatus and stimulus materials

Testing was conducted in a computer-controlled (Cromenco Z-2D) operant chamber. Stimuli were rear projected onto a 12 cm \times 18 cm translucent projection screen using a slide projector (Kodak 760H). This screen was located 23 cm behind a clear Plexiglas 9 cm \times 9.3 cm response key centrally located in the front wall of the chamber. Stimulus duration was controlled by a computer-controlled shutter located in front of the projector lens. A food hopper was centrally located 7.5 cm below the response key and delivered mixed grain. A house light was located in the ceiling and was illuminated at all times.

All exemplars for both categories were photocopied from two edited collections of animal line drawings (Harter, 1979; Iyari, 1979). These items were then photographed with a SLR camera (Nikon FE2, 1/30th of a second at F 5.6, Kodachrome 64 slide film) and presented as 35 mm slides. A blue filter (Tiffen 80B) was used to compensate for the incandescent lighting used to illuminate the images during photography. The stimulus manipulations described below were performed on photocopied images prior to being photographed. When projected, the stimuli subtended a vertical visual angle of approximately 25°.

2.3. Procedure

2.3.1. Baseline categorization

Throughout all tests, the pigeons were maintained on a baseline categorization discrimination consisting of daily session (40 S+ and 40 S- presentations). The bird category was designated the S+ category and the mammal category designated the S- category for all five pigeons. Each trial consisted of a randomly determined category exemplar presented for 30 s. Pigeons were rewarded for pecks to the S+ category on a variable interval schedule (VI-45) that resulted in reinforcement on 66% of trials on average. Pecks to the S- category were not reinforced (i.e., extinction). Sixteen different orders of slide presentations were tested and employed different combinations of forward and backward projector motion to prevent any sequential memorization. Between tests, these baseline items were irregularly changed to incorporate new exemplars of each category to limit item memorization. Trials were separated by variable inter-trial intervals of 3-15 s that were independent of the randomized forward and backward positioning of the slide tray between trials.

2.3.2. Stimulus analytic test sessions

The general procedures were essentially identical for each of the tests. The primary variable was the nature of the stimulus manipulation tested, with minor adjustments in the details of session organization depending on the number of conditions and exemplars tested. Each test session consisted of 80 trials. The first 20 trials

were used as a warm-up period during which only baseline category stimuli were shown. Test stimuli were randomly placed into the remaining 60 trials with the constraint that at least two baseline trials occur between test trials. Test images replaced randomly selected baseline images in the slide tray. All temporal parameters were identical to the baseline sessions. These tests were conducted as *probe trials* during which no reinforcement was allowed. These probe trials allowed for the uncontaminated measurement of peck rate without the presence of food presentations. The rationale and details of the different analytic tests are described next.

3. Specific test procedures and results

3.1. Test 1: analysis of figure/ground – procedures

This test examined the contribution of the animal figure relative to its surrounding background. Because the original images were naturalistic in origin, there were correlations between the animal category and its setting. For each session, a set of category exemplars were modified, choosing baseline images with visible background characteristics. In the *figure* condition, the entire background was removed, leaving just the solitary animal. In the *Ground* condition, the animal figure was deleted, leaving just the background. Illustrative examples of each of the conditions are displayed in Fig. 1. In total, twenty exemplars (10 birds/10 mammals) were tested in these figure-only and background-only conditions. Each session tested two bird and mammal exemplars in each condition as unreinforced probe trials. Over 10 sessions, each of these exemplars was tested twice in each condition.

To test the role of *figure* vs. *ground* in categorical transfer performance, the pigeons were tested with novel exemplars. In total, 16 novel exemplars (8 birds/8 mammals) were tested in the *figure*, *ground* and a *complete image* transfer (neither property deleted) conditions. Each session tested one novel bird and mammal exemplar in each of these three conditions once as an unreinforced probe trial. Testing was conducted for a total of eight sessions.

3.2. Test 1: analysis of figure/ground - results

The left panel of Fig. 2 shows the mean peck rate to the baseline, figure, and ground conditions testing familiar exemplars. The baseline categorization continued to be well discriminated as indicated by the higher peck rates by all five pigeons to exemplars of the bird category (black bars) compared to the mammal category (gray bars). The figure condition supported virtually equivalent levels of discrimination. In contrast, the ground condition supported a much lower level of discrimination than either of these conditions. As to be expected from this pattern, a repeated measures ANOVA (Category × Condition) confirmed a significant interaction between category and condition, F(2,8) = 31.1 (an alpha of p < .05 was used to judge the significance of statistical tests). Separate two-tailed paired *t*-tests (df = 4) comparing peck rates to each category within each condition confirmed the existence of significant categorical discrimination in the baseline and *figure* conditions (ts > 2.76). For the ground condition, there was little or no evidence for discrimination among the categories, t(4) = 2.4, p = .07.

The right panel of Fig. 2 shows the mean peck rate to the baseline, complete, figure, and ground conditions in the test with novel exemplars. Again, the baseline categorization continued to be well discriminated. The novel figure condition supported similar levels of discrimination among the categories. The complete figure transfer condition supported transfer, but not to the same degree as the baseline or novel figure conditions (see below). The novel ground condition supported no discrimination, with essentially equivalent peck rates to each category. Again, a repeated



Fig. 1. Illustrative examples of the baseline, figure, and ground conditions of both categories used in Test 1.

measures ANOVA (Category × Condition) confirmed there was a significant interaction between category member and condition, F(3, 12) = 11.7. Separate paired *t*-tests comparing categorical discrimination in each condition revealed significant differences in peck rate to each category for the baseline, t(4)=5.3 and novel figure conditions t(4)=7.1. The moderate transfer results for the complete figure condition were due to one pigeon performing at near chance in this specific condition. With this bird included, the differences between the categories were not significant, t(4)=2.2, p < .09. When excluded, the other four pigeons did show significant discrimination transfer, t(3)=3.4. All five pigeons performed poorly in the novel ground-only condition and there were no significant differences among the categories for this condition.

Thus, the pigeons' discrimination of the categories was almost exclusively controlled by the animal figure in both tests. Some pigeons may have learned a little something about the item-specific background features associated with familiar bird and mammal exemplars, but this learning did not generalize to novel images where features of the background potentially correlated with each category (e.g., trees versus open ground) were not sufficient. In fact, the presence of the background cues may have impeded generalization to novel animal figures as suggested by the better transfer performance observed in the novel figure condition in which they were absent. The next test tried to understand better what parts of the animal were most important.

3.3. Test 2A: analysis of figural components (head and body chimeras) – procedures

This test examined the contribution of the animal figure's body vs. its head to the discrimination. This was done by testing chimera animals created from conflicting information from both categories by combining the head of an animal from one category with the body of an animal from the other category and control animals combining heads and bodies from different exemplars of the same category. Four chimera conditions were created. These consisted of *control bird/bird* exemplars made from the head and body of two different birds, *control mammal/mammal* exemplars combined from two different mammals, *bird/mammal chimeras* (head/body) combining the conflicting head of a bird with the body of a mammal, and *mammal/bird chimeras* combining the conflicting head of a a mammal with the body of a bird. In making the chimeras, the head and body components of eighteen different animals were used. Care was taken in combining these parts to make creatures with



Fig. 2. Mean peck rate to the different test conditions examined in Test 1. The left panel shows these results for tests involving familiar exemplars. The right panel shows the results for tests involving novel exemplars. Error bars represent the SEM of each condition.



Fig. 3. Illustrative examples of the different chimera test conditions for combinations of both categories examined in Test 2A.

smooth contours and reasonable proportions. The background was also removed for these and future test stimuli. Illustrative examples of the conditions are displayed in Fig. 3. Each session tested two examples each of the four conditions. The control bird/bird and mammal/mammal test trials were differentially reinforced as they contained all the properties of the categories they represented. The two chimera conditions were tested as non-reinforced probe trials. Nine test sessions were conducted.

To test the role of these properties in determining categorical transfer performance, the pigeons were next tested with novel chimera exemplars. Thus, the only difference from the previous head vs. body test was that the chimera stimuli were made from parts of birds and mammals to which the pigeons had not previously been exposed. Each session tested novel exemplars in each of the four conditions twice. All test trials randomly appeared within a session and were tested as non-reinforced probes, including the control conditions. This test was conducted for six sessions, at which point the test images were reflected to face the opposite direction and another similar six-session test block conducted. Reflecting the images was done to minimize the effects of the prior exposure and maintain its novelty.

3.4. Test 2A: analysis of figural components (head and body chimeras) – results

The left panel of Fig. 4 shows the mean peck rates to the baseline and familiar chimera test conditions. The baseline categorization continued to be significantly discriminated and provides a good reference point for interpreting the birds' reactions to the chimeras. The pigeons exhibited excellent discrimination to those chimeras where the head and the body came from the same category. The peck rate in the bird/bird chimera condition was much greater and significantly different from that in the mammal/mammal chimera condition, t(4) = 7.1. In fact, this chimera discrimination was equivalent to that of the baseline condition as peck rates in the bird/bird chimera condition were not significantly different from the baseline bird category. Likewise, peck rates in the mammal/mammal chimera condition were found not to be significantly different than in the baseline mammal condition. Together, these results indicate that recombining heads and bodies of animals from within the same category did not alter the discrimination.

Results from the conflict chimera condition indicated that features associated with the body, and not the head, controlled the discrimination. Here the peck rate to the mammal-head/bird-body condition was not significantly different from that of the baseline bird condition while being significantly different from the baseline mammal condition, t(4) = 9.0. Correspondingly the peck rate to the bird-head/mammal-body condition was significantly different from the baseline bird condition, t(4) = 7.2, and even better discriminated than the baseline mammal condition, t(4) = 2.9. Finally, discrimination of the two consistent chimera conditions was not any better than that observed between the two conflict conditions. Overall, the results indicate that the head made little contribution to the discrimination.

The right panel of Fig. 4 shows the mean peck rates to the baseline and novel chimera test conditions. For the novel chimeras that maintained a consistent categorical structure, the pigeons' discrimination continued to be excellent, although slightly reduced from baseline levels. Peck rates to the novel bird/bird chimera condition were significantly different from those of the novel mammal/mammal chimera condition, t(4)=12.9. Peck rates to the novel bird/bird condition were not significantly different



Fig. 4. Mean peck rate to the different test conditions examined in Test 2A. The left panel shows these results for tests involving familiar exemplars. The right panel shows the results for tests involving novel exemplars. Error bars represent the SEM of each condition.

relative to the baseline bird condition, but peck rates to the novel mammal/mammal condition were not suppressed to an equivalent degree in comparison to the baseline mammal condition, t(4)=5.3. This difference suggests that the novelty of the test items was detected.

Results from the novel conflict chimera condition again supported the conclusion that the body, and not the head, controlled the discrimination. Here peck rates to the novel mammalhead/bird-body condition were not significantly different from that of the baseline bird condition, but were significantly different from the baseline mammal condition, t(4) = 12.2. This result indicates that when categorical features within the stimuli were in conflict, it was the body that prevailed and ultimately determined the classification. Consistent with these results were the peck rates to the novel bird-head,/mammal-body condition. In this condition, the peck rates were significantly different from the baseline bird condition, t(4) = 7.2, and only slightly, but significantly, higher than the baseline mammal condition, t(4) = 3.2.

Both of these tests clearly indicate that the pigeons were consistently classifying the chimera conflict stimuli based on properties of the body, and not the head, regardless of their familiarity with the stimuli. Our results converge with Ghosh et al.'s (2004) finding with pigeons that body-associated cues were more important in a similar chimera test examining the categorical discrimination of cats and dogs by pigeons. The next test explored some properties of that might be responsible for this control by the "body."

3.5. Test 2B: analysis of figural components (number and type of legs) – procedures

One major difference between mammals and birds concerns their typical number of legs. The next test examined the contribution of the number of the exemplar's limbs to the categorical discrimination. This was done by testing chimera animals with varying number of legs. One manipulation involved making the number of legs match those of the other category. In the quadrupedal bird condition, a set of birds had a matching set of their legs added to the front part of their bodies to give them four legs. In the bipedal mammal condition, a set of mammals had their front legs removed to make them have two legs (somewhat similar to birds). The second manipulation involved making chimera animals that had the number and type of legs from the other category. In the bird/mammal (body and head/legs) condition, a set of birds had a matched set of legs from a mammal placed on their bodies. In the mammal/bird (body and head/legs) condition, a set of mammals had a matched set of legs from a bird placed on their bodies. Again, in constructing these stimuli, care was taken to combine these parts in a proportional manner with smooth contours (although this was more difficult than for the head/body manipulation). Illustrative examples of the conditions are shown in Fig. 5. Each session tested two examples of each of the four conditions, plus two trials that tested the original baseline figures (five from each category) used to construct the stimuli. All 12 of these tests randomly appeared within a session and were tested as non-reinforced probe trials. Ten test sessions were conducted.

To test the role of these properties in determining categorical transfer performance, the pigeons were again tested with novel exemplars of these conditions. Thus, the only difference from the previous test was that the stimuli were made from novel bird and mammals parts. Each session tested novel exemplars in each of the four conditions twice, plus two trials that tested the unaltered novel bird and mammal exemplars (six from each category) used to construct the stimuli. All 12 test trials randomly appeared within a session and were tested as non-reinforced probes. Six different sets of these conditions were tested. This testing was conducted for



Fig. 5. Illustrative examples of the different chimera leg conditions for both categories examined in Test 2B.

12 sessions, with each set tested twice (the second time right/left reflected).

3.6. Test 2B: analysis of figural components (number and type of legs) – results

Overall, the number of legs present on the body made little substantive contribution to the pigeons' categorical discrimination. The left panel of Fig. 6 shows the mean peck rates to the baseline and test conditions. When the bird-bodied test figures with either four bird (B/B) or four mammal legs (B/M) were compared with paired t-tests (dfs = 4) to the baseline bird discrimination there were no significant differences in peck rates, indicative of their classification as bird-like. Similarly when the mammal-bodied test figures with either two mammal (M/M) or two bird legs (M/B) were compared to the baseline mammal discrimination there were no significant differences in peck rates among these conditions, indicative of their classification as mammal-like. The peck rates for the consistent mixed body and legs conditions (B/B and M/M) suggested slightly better control than when inconsistent chimera mixtures were used (B/M and M/B), but statistical comparisons found this to be unreliable.

The right panel of Fig. 6 shows the mean peck rates to the transfer and test conditions for tests involving novel figures. Again the pigeons showed significant transfer when tested with novel figures of each category, t(4) = 4.4. When the bird-bodied test figures with either four bird (B/B) or four mammal legs (B/M) attached to them were compared with paired t-tests to the transfer bird discrimination there were no significant differences in peck rates, indicating the classification of these novel figures as bird-like. Comparison of the overall levels of responding to the novel mammal-body test figures with either two mammal (M/M) or two bird legs (M/B) attached to them were consistent with their classification as mammal-like. However, when compared with paired *t*-tests to the transfer mammal discrimination there were significant differences in peck rates for both the M/M, t(4) = 5.2, and M/B conditions, t(4) = 2.8. However, both of these conditions supported significantly lower peck rates than either the B/B or M/B four legs conditions, all ts(4) > 2.8. This observation, in addition to the generally lower peck rates to two legged figures, indicate their classification of these modified novel figures as mammal-like. Thus, both tests consistently suggest that the number and type of legs attached to the body are not the properties responsible for the present categorical discrimination.

3.7. Test 3A: analysis of global and local features (figural separation) – procedures

The next tests examined the contribution of global form and local features of the animal to the discrimination. The first test



Fig. 6. Mean peck rate to the different test conditions examined in Test 2B. The left panel shows these results for tests involving familiar exemplars. The right panel shows the results for tests involving novel exemplars. Error bars represent the SEM of each condition.

examined how separating and scrambling different portions of the exemplars influenced discrimination. The global form of nine birds and nine mammal figures with their backgrounds removed were separated in three approximately equivalent segments containing the head, torso, and rear of each animal. From these parts four different conditions were created. These consisted of the separated condition, where the parts were separated by approximately 2 cm from each other along the animal's principal and canonical orientation, and the close and scrambled condition in which the three parts were scrambled, but placed adjacently to each other. In the *separated and scrambled* condition the head and rear parts were exchanged while in the random and scrambled condition the three parts were positioned randomly with the restriction that the normal order could not appear. Illustrative examples of these conditions are displayed in Fig. 7. Each session tested two baseline exemplars (one bird/one mammal) in each of the four conditions. All eight test trials randomly appeared within a session and were tested as non-reinforced probes. Nine test sessions were conducted.

Following this, the pigeons were tested with novel exemplars. Each session tested novel exemplars in each of the four conditions twice, plus two trials that tested the novel bird and mammal exemplars used to construct the stimuli. All ten test trials randomly appeared within a session and were conducted as non-reinforced probes. Six different sets of these conditions were tested. This testing was conducted for 12 sessions, with each set tested twice (the second time right/left reflected).

3.8. Test 3A: analysis of global and local features (figural separation) – results

The results for this test are slightly more complex than those from earlier tests in that the outcomes of the manipulation appeared to depend on familiarity of the figures. The left panel of Fig. 8 shows the mean peck rates to the baseline and test conditions with the familiar stimuli. Here the global organization of the body parts had little effect on discrimination of familiar stimuli of each category. The baseline categorization continued to be clearly and significantly discriminated, t(4)=5.1. The same was also true for all four test conditions as the bird category supported significantly higher peck rates than the mammal category regardless of its spatial organization, ts(4)>2.7. When compared with baseline responding and across tests of items within the same category, there were no significant differences among the conditions confirming that peck rates were equivalent across conditions. Thus, independent of whether the body region of the animal had been separated, scrambled, or both, there was no effect on the discrimination of highly familiar items of bird/mammal categories.

The right panel of Fig. 8 shows the mean peck rates for the transfer and test conditions with the novel exemplars. These averages are computed based on four of the five pigeons that were tested, because one pigeon did not transfer to the novel figures in this test and pecked at very low rates to all of these transfer stimuli. For the other four pigeons, the transfer stimuli of complete figures supported significant categorical discrimination, t(3) = 3.9, again. Spatial separation of body segments reduced discrimination between categories as peck rates decreased with birds and increased with mammals. Although responding to the bird category was numerically higher, there was no significant differences in peck rates between the two categories for either separated test condition, ts(3) < 2.1. When the same figures were compacted close together, however, significant differences between the two categories emerged with peck rates to novel bird conditions greater than mammal conditions, ts(3)>4.1. This was true regardless of how the position of the segments were scrambled within this compact figure.

This pattern of results suggests that the global configuration of the animal figures was more important than the specific arrangement of the local parts, especially for novel stimuli. Across all conditions, scrambling the local segments did not impact the discrimination when compared to unscrambled stimuli. When the segments were more widely separated, however, the categorical



Fig. 7. Illustrative examples of the different test conditions for both categories examined in Test 3A.

discrimination of the novel stimuli was reduced. This result implies that pigeons had more difficulty recognizing the critical categorical cues, except when the segments were close together. In contrast, the results of the first test with the familiar items suggested that any such spatial and organizational factors were less important, perhaps because other categorical and item-specific cues had been well encoded and were still available despite the different spatial manipulations.



Fig. 8. Mean peck rate to the different test conditions examined in Test 3A. The left panel shows these results for tests involving familiar exemplars. The right panel shows the results for tests involving novel exemplars. Error bars represent the SEM of each condition.



Fig. 9. Illustrative examples of the different brightness conditions for both categories used in Test 3B.

3.9. Test 3B: analysis of global and local features (figural uniformity) – procedures

The next test examined how eliminating the local interior detail within stimuli affected performance. The interior detail of eight familiar birds and eight familiar mammals (backgrounds removed) were replaced by uniform gray-scaled textures. The texture patterns of 100% 70%, 50%, 30%, and 0% gray scale levels were used to vary the amount of contrast in the image. Five different trial types ranging from black silhouette to outlined contour were created. Illustrative examples of the conditions are displayed in Fig. 9. Each session tested two baseline exemplars (one bird/one mammal) in each of the five conditions. All 10 test trials randomly appeared within a session. Twenty-four test sessions of this type were conducted. Unlike the other tests, these test trials were differentially reinforced as we were concerned that their distinctive nature might cause the pigeons to learn that stimuli without local details never yielded reward. No test with novel exemplars was conducted because of their poor performance with the familiar exemplars.

3.10. Test 3B: analysis of global and local features (figural uniformity) – results

Eliminating the interior featural detail by making the animal figures uniform in appearance produced by far the most mixed set of reactions from the pigeons among the different tests. The right panel of Fig. 10 shows the mean peck rates for the transfer and test conditions for all five pigeons. The summed test results for all



Fig. 10. Mean peck rate to the different brightness conditions examined in Test 3B. Error bars represent the SEM of each condition.

five subjects suggest that only with intermediate gray levels was their categorical discrimination maintained to some degree. The pattern across conditions in Fig. 10 is due to the impact of only two pigeons, however, as the three remaining subjects showed no effect of interior brightness. Both of these pigeons were able to perform a diminished, but significant, discrimination of figures with gray stimuli of intermediate to black values (silhouette). A repeated measures ANOVA (Brightness level × Category) on peck rates for these two birds confirmed the presence of an interaction between these two factors, F(1,4) = 11.9. For both of these pigeons, there were significant differences between bird and mammal conditions at the 70%, and 100% levels across sessions, ts(23)>2.5. Of these two, one pigeon could also discriminate at the 50% level, t(23) > 2.3, and the other was very close, t(23) = 2, p = .07. Neither of these two pigeons could discriminate categories at the 30% level or the 0% outline figures, exhibiting equivalent peck rates for each category. These results indicate that these two pigeons needed a certain level of brightness that approximately matched or exceeded the average brightness of the original stimuli to perform the discrimination accurately.

The other three birds performed differently. The pigeon that performed poorly in the prior test with separated and scrambled body parts, again responded at very low rates to all of these test stimuli in both categories. This result suggests that the elimination of the interior local detail strongly impacted this pigeon's ability to recognize them as reinforced stimuli. The remaining two pigeons exhibited more typical levels of responding to these stimuli, but also showed no discrimination of the categorical test stimuli at any of the five gray-scale levels. A repeated measures ANOVA (Brightness level × Category) on peck rates for these three birds revealed no significant interaction between these two factors, unlike the first two pigeons, or significant main effects of either Category or Brightness level. The absence of any discrimination of the categories by these three pigeons suggests that the interior local details or black and white textural variation within the animal figures were part of their representation of the categories. When removed and made uniform, these pigeons could no longer discriminate the figures based on the silhouette contour of the familiar training stimuli, regardless of its relative brightness.

3.11. Test 4A: analysis of figural orientation – procedure

The final test examined the role of stimulus orientation to the discrimination. This was done by reorienting the principal axes of the categories to different degrees of rotations. Three conditions were tested. These consisted of *reversed to other category* exemplars made from rotating birds 45° down to a horizontal orientation typical of most mammals and rotating mammals 45° up to the diagonal


Fig. 11. Illustrative examples of the different orientation conditions for both categories examined in Test 4.

orientation typical of birds; 90° condition where the principal axes of the animals from each category were at 90° relative to horizon, and the 135° condition where the principal axes of the animals from each category were rotated to be 135° relative to its original axis. Illustrative examples of each of the conditions are displayed in Fig. 11. Ten test sessions were conducted, with each testing two familiar birds and mammals in the different orientation conditions. All test trials randomly appeared within a session and were tested as non-reinforced probes. Finally, the pigeons were again tested with novel exemplars. Six test sessions were conducted, with each testing two novel birds and novel mammals in the different orientation conditions. All test trials randomly appeared within a session and were tested as non-reinforced probes.

3.12. Test 4: analysis of figural orientation - results

Reorienting the animal figures had an effect on the pigeons' ability to perform their categorical discrimination with both familiar and novel stimuli. The left panel of Fig. 12 shows the mean peck rates for the baseline and orientation conditions for all five birds. For the two conditions that rotated the familiar stimuli to 90° or 135°, this manipulation effectively eliminated the pigeons' discrimination of the familiar stimuli, indicating that their orientation was critical. The reversed orientation condition produced mixed results with familiar stimuli, with two pigeons able to partially maintain the original discrimination (the reason for the difference in the figure), while the remaining three pigeons dropped to chance levels of discrimination. Not surprisingly, a repeated measures ANOVA (Orientation [Baseline and Tests] × Category) on peck rates for all birds confirmed the significant interaction of Orientation and Category, F(3, 12) = 5.0. Subsequent paired *t*-tests among the different conditions confirmed the significant difference in peck rate between the categories for the baseline condition, t(4) = 3.1, but not for either the reversed, 90° and 135° conditions where orientations were altered.

Reorienting novel animal figures was equally effective as disrupting the discrimination, as all five birds now had difficulty with those figures during novel transfer testing. The right panel of Fig. 12 shows the mean peck rates for the transfer and orientation



Fig. 12. Mean peck rate to the different orientation conditions examined in Test 4. The left panel shows these results for tests involving familiar exemplars. The right panel shows the results for tests involving novel exemplars. Error bars represent the SEM of each condition.

conditions for all five birds. Baseline performance continued to be good, as was transfer performance to the novel exemplars of each category. Changing the orientations of these novel figures, however, created problems as evidenced by the poor discrimination across the three orientation test conditions. A repeated measures ANOVA (Orientation [Transfer and Tests] × Category) on peck rates for all birds confirmed the existence of a significant interaction of Orientation and Category, F(3, 12) = 3.9. Subsequent paired *t*-tests among the different conditions confirmed a significant difference in peck rate between categories for the transfer condition, t(4) = 5.2, but not among any of the three conditions where orientation was altered.

4. General discussion

The current results paint one of the most complete descriptions yet of the visual factors controlling a categorical discrimination by pigeons. Our pigeons exhibited the major hallmarks used to identify and invoke accounts involving categorical representation. They were able to accurately discriminate large numbers of polymorphous exemplars from two open-ended visual categories and show significant and consistent transfer across multiple tests with novel exemplars of these categories. From the totality of the results from the different stimulus manipulations, it appears that these two categories were visually discriminated by the pigeons by segregating the textured animal figures from the background and extracting the orientation of the principal axis of the animal's body.

Several outcomes suggested that bodily orientation was critical to the categorical discrimination. The most direct evidence was the results from the final orientation test, where reorienting the animal figures disrupted the discrimination. The chimera tests that recombined the categories with either conflicting heads or legs and showed that body (and in its typical orientation) – not heads or legs – controlled the discrimination (see also Ghosh et al., 2004). The figure/ground test suggested that transfer performance might even be better when the background was removed. Eliminating the background's clutter and noise may have made orientation of the animal more prominent.

Tests with the body divided, scrambled, and spatially separated, also suggested a body-related computation, like orientation, was critical. Scrambling the body parts had little effect on performance, most likely because this manipulation did not change the basic orientation of the figure. However, when these same body-part manipulations were made to novel bodies, the discrimination suffered, suggesting that familiarity was necessary to recognize these parts across their separation. Taken together, these different results suggest a conclusion that the principal body axis was the primary cue underlying their categorical discrimination and transfer.

Analyses looking at the orientation of the baseline exemplars additionally confirmed a role for bodily orientation. For the 80 baseline exemplars tested during the extended number of sessions in Test 3A, the average orientation of the bird exemplars was 44° from the horizontal axis of the picture, while the mammal exemplars averaged 17° . Looking at overall peck rates for all items as a function of orientation revealed a significant correlation across categories (r(78) = .51) but not within categories (bird: r(38) = .17; mammal: r(38) = .12), possibly due to within-category orientation similarity.

The results also indicate that cues other than orientation had some effect on the pigeons' processing and representation of these stimuli. First, the pigeons were clearly performing some kind of figure/ground segregation (Cook and Hagmann, 2012; Lazareva et al., 2006a). The various and complex backgrounds present in these stimuli contributed little to the discrimination and perhaps even interfered with it because they masked the figural orientation of the animals. Because of their original intention as public illustrations, the vast majority of the images were in landscape format, making the orientation of the entire image useless to category identification. This suggests the pigeons were extracting the "object" from the illustration perhaps by identifying the most central, enclosed, dense pattern in the stimulus. Second, the pigeons' discrimination clearly depended upon the 'fill' or interior features of the body figure. We showed that outlines of the animal figures alone were not sufficient to maintain the discrimination, despite orientation being a readily available cue in these cases. Similar failures to find transfer from shaded to outlined figures have been found with pigeons in other contexts (Aust and Huber, 2002; Cabe and Healey, 1979; Peissig et al., 2005). In our experiments, when the interior detail or texture of the figures was removed and replaced with uniform areas of different brightness, the discrimination of categories did suffer for the majority of the birds. Moreover, reorienting the figures to the other category's typical orientation produced a general disruption in performance (as opposed to misclassification), suggesting that more than simple body orientation was involved in these classifications by pigeons.

Together the results suggest that the pigeons' representation of categories depended upon having a certain absolute level of texture, detail, or brightness being present in the interior of the figure. This would be consistent with the idea that surface information may also more important in defining objects for these animals (cf. Cook et al., 2012; Loidolt et al., 1997; Peissig et al., 2005). Finally, item-specific memorization also played a role similar to that found in several other experiments (e.g., Aust and Huber, 2001). In several tests, discriminations of the familiar figures survived stimulus manipulations, such as figural separation, that interfered with discrimination of novel items. This suggests that there were miscellaneous, and likely idiosyncratic, item-specific features that were memorized about the familiar stimuli over and above the categorical factors that mediated novel transfer.

While the evidence that pigeons can solve a wide variety of complex pictorial classifications is bountiful, the results of the current study raise important questions about the meaning of such results. Part of the initial excitement and importance attributed to such findings was the implication that human-like concept learning was a cognitive ability shared widely in the animal kingdom. Yet, the body of evidence in the current study suggests the hypothesis that the pigeons reduced the apparent complexity of this categorical discrimination to a simple set of a few critical visual features. While the nature of the controlling properties has been examined for only a handful of categories with different species, the results are similar in showing that the complex categorical discrimination seems to be reduced to a simple set of critical visual features (e.g. Aust and Huber, 2002; Brooks et al., in press; D'Amato and Van Sant, 1988; Lazareva et al., 2006b; Troje et al., 1999). If such cases are representative of the general processing of pictorial stimuli by nonhuman animals, then one could question whether we been truly studying concept learning with this preparation over the last 50 years. The answer to this question has numerous implications.

Disregarding for the moment its relations to human concept learning, the results of this field-wide research effort have been very important in revealing the nature of the discrimination learning process, especially in pigeons. What the evidence clearly indicates is that pigeons, and potentially other animals, are exquisitely sensitive to discovering and extracting relevant information from highly complex multidimensional displays. No matter how complex and polymorphous the categories are (e.g., abstract art, Watanabe et al., 1995), if the pictures contain features correlated with the categories, then pigeons appear quite capable of finding, isolating, and using them. Whether or not they have conceptualized such solutions, there is certainly no doubt that their visual/cognitive equipment is second to none when it comes to feature discovery and identification. For computers, the latter is a particular difficult problem (Rumelhart and Zipser, 1985). It is only when challenged by categories lacking such coherent featural similarity that we see a breakdown in this exceptional feature extraction ability in pigeons, such as in the case of discriminating living versus non-living items (Roberts and Mazmanian, 1988).

This perceptual/cognitive reduction of the stimulus space to a few dimensions is seemingly characteristic of this species' processing of complex pictorially-based categorization problems. One of the chief difficulties of past experiments using compound stimuli having multiple dimensions is getting the birds to move away from this approach. Several experiments had found that pigeons have a difficult time learning artificial polymorphous concepts that require their simultaneous processing of multiple dimensions of the stimuli for their solution (Lea et al., 1993, 2006b). This is not to say they cannot attend to multiple features, but they seem generally resistant to doing so. When such featural solutions fail, pigeons can always fall back on their considerable capacity to memorize patterns and configurations (Cook and Fagot, 2009; Cook et al., 2005; Fagot and Cook, 2006; Vaughan and Greene, 1984), which gives the birds the capacity to acquire pseudo-categories, and likely many other types of complex discriminations.

This pattern suggests that pigeons have two sources for information during learning; one tied to features correlated with class membership, and the other tied to the features of each item. The category-related and item-specific sources are both valuable and it is their balance and relative competition that determines the rate of learning and the eventual nature of the representation used as the basis for the discrimination. These sources may lead to a cascade of apparent strategies used by pigeons when faced with solving complex discriminations. If item-specific information is deemphasized by having large number of similar items grouped together, then categorical-like behavior emerges, likely based on a reduced dimensionality of the stimuli. On the other hand, if the processing of such absolute information is demanded, by say the randomized assignment of stimuli to responses, then the pigeons memorize large amounts of item-specific information. Of course, both can be made part of the same experiment and you see both sources in action. Using artificial multidimensional stimuli, Cook and Smith (2006) found that dimensional abstraction seemed to precede item memorization. That pigeons also learn consistent categories relations faster than pseudo-categorizes further suggests that the memorization of items is secondary or emerges more slowly than abstraction-based analysis. The duality and competitive nature of these factors in discrimination learning has been captured in a variety of models that pivot around these distinctions (Anderson and Betz, 2001; Nosofsky et al., 1994; Soto and Wasserman, 2010). Although the computational mechanisms are different, these models attempt to capture the constant interplay between these two sources of information. Understanding better how pigeons rapidly discover and identify the key features in complex pictures, along with identifying the conditions that allocate control between common and item-specific features, remain important topics for investigation.

The question of whether we have been investigating categorization in animals remains unanswered. At one level, the answer is clearly yes. The last several decades of research has substantially advanced and expanded our understanding of how discrimination learning operates and revealed the remarkable ability of pigeons to process complex stimulus situations. The natural behavior of this animal on the street would unlikely lead one to suspect that its small brain contained such remarkable and powerful computational resources. Beyond their superficial similarities, however, it is not clear yet that the mechanisms involved are comparable to those responsible for conceptual behavior in humans. Two possibilities seem likely. The first is that with more detailed examinations, we will establish that human conceptual abilities operate in a way different from those in pigeons. A second possibility is that we have perhaps overestimated the sophistication of human thought on this issue, and that at their core, the process of feature discovery and classification operate very similarly at least when challenged to processing complex pictorial information of varying visual similarity (e.g., Gluck and Bower, 1988).

Acknowledgments

The conduct and preparation of these experiments were supported by grants from the National Science Foundation and the National Eye Institute. The authors thank Justin Sayde, Melissa Langer, and Muhammad Qadri for helpful comments on this paper.

References

- Anderson, J.R., Betz, J., 2001. A hybrid model of categorization. Psychon. Bull. Rev. 8, 629–647.
- Aust, U., Huber, L., 2001. The role of item- and category-specific information in the discrimination of people- versus nonpeople images by pigeons. Anim. Learn. Behav. 29, 107–119.
- Aust, U., Huber, L., 2002. Target-defining features in a people-present/people-absent discrimination task by pigeons. Learn. Behav. 30, 165–176.
- Aust, U., Huber, L., 2003. Elemental versus configural perception in a peoplepresent/people-absent discrimination task by pigeons. Learn. Behav. 31, 213–224.
- Brooks, D.I., Ng, K.H., Buss, E.W., Marshall, A.T., Freeman, J.H., Wasserman, E.A. Categorization of photographic images by rats using shape-based image dimensions. J. Exp. Psychol.: Anim. Behav. Process., in press.
- Cabe, P.A., Healey, M.L., 1979. Figure-background color differences and transfer of discrimination from objects to line drawings with pigeons. Bull. Psychon. Soc. 13, 124–126.
- Cerella, J., 1986. Pigeons and perceptrons. Pattern Recogn. 19, 431-438.
- Cook, R.G., 1993. The experimental analysis of cognition in animals. Psychol. Sci. 4, 174–178.
- Cook, R.G., Fagot, J., 2009. First trial rewards promote 1-trial learning and prolonged memory in pigeon and baboon. Proc. Natl. Acad. Sci. U. S. A. 106, 9530–9533.
- Cook, R.G., Hagmann, C.E., 2012. Grouping and early visual processing in avian vision. In: Lazareva, O.F., Shimizu, T., Wasserman, E.A. (Eds.), How Animals See the World: Behavior, Biology, and Evolution of Vision. Oxford University Press, London.
- Cook, R.G., Smith, J.D., 2006. Stages of abstraction and exemplar memorization in pigeon category learning. Psychol. Sci. 17, 1059–1067.
- Cook, R.G., Wright, A.A., Kendrick, D.F., 1990. Visual categorization by pigeons. In: Commons, M.L., Herrnstein, R., Kosslyn, S.M., Mumford, D.B. (Eds.), Quantitative Analyses of Behavior: Behavioral Approaches to Pattern Recognition and Concept Formation. Lawrence Erlbaum, Hillsdale, NJ, pp. 187–214.
- Cook, R.G., Levison, D.G., Gillett, S.R., Blaisdell, A.P., 2005. Capacity and limits of associative memory in pigeons. Psychon. Bull. Rev. 12, 350–358.
- Cook, R.G., Qadri, M., Kieres, A., Commons-Miller, N., 2012. Shape from shading in pigeons. Cognition 124, 284–303.
- D'Amato, M.R., Van Sant, P., 1988. The person concept in monkeys (*Cebus apella*). J. Exp. Psychol. Anim. Behav. Process. 14, 43–55.
- Dittrich, L., Rose, J., Buschmann, J.U.F., Bourdonnais, M., Güntürkün, O., 2010. Peck tracking: a method for localizing critical features within complex pictures for pigeons. Anim. Cogn. 13, 133–143.
- Fagot, J., Cook, R.G., 2006. Evidence for large long-term memory capacities in baboons and pigeons and its implication for learning and the evolution of cognition. Proc. Natl. Acad. Sci. U. S. A. 103, 17564–17567.
- Ghosh, N., Lea, S.E.G., Noury, M., 2004. Transfer to intermediate forms following concept discrimination by pigeons: chimeras and morphs. J. Exp. Anal. Behav. 82, 125.
- Gluck, M.A., Bower, G.H., 1988. From conditioning to category learning: an adaptive network model. J. Exp. Psychol. Gen. 117, 227–247.
- Greene, S.L., 1983. Feature memorization in pigeon concept formation. In: Commons, M.L., Herrnstein, R.J., Wagner, A.R. (Eds.), Discrimination Processes, vol. 4. Ballinger, Cambridge, MA, pp. 209–229.

Harter, J., 1979. Animals. Dover Publications, New York.

- Herrnstein, R.J., 1979. Acquisition, generalization, and discrimination reversal of a natural concept. J. Exp. Psychol. Anim. Behav. Process. 5, 116–129.
- Herrnstein, R.J., Loveland, D.H., 1964. Complex visual concept in the pigeon. Science 146, 549–551.
- Huber, L., 2001. Visual categorization in pigeons. In: Cook, R.G. (Ed.), Avian Visual Cognition. (On-line).
- Iyari, H., 1979. Book of Animals. Mayflower Books, Clinton, NJ.
- Lazareva, O.F., Castro, L., Vecera, S.P., Wasserman, E.A., 2006a. Figure-ground assignment in pigeons: evidence for a figural benefit. Percept. Psychophys. 68, 711–724.

- Lazareva, O.F., Freiburger, K.L., Wasserman, E.A., 2006b. Effects of stimulus manipulations on visual categorization in pigeons. Behav. Process. 72, 224–233.
- Lea, S.E.G., Lohmann, A., Ryan, C.M.E., 1993. Discrimination of five-dimensional stimuli by pigeons: limitations of feature analysis. Quart. J. Exp. Psychol. 46, 19–42.
- Lea, S.E.G., Goto, K., Osthaus, B., Ryan, C.M.E., 2006a. The logic of the stimulus. Anim. Cogn. 9, 247–256.
- Lea, S.E.G., Wills, A.J., Ryan, C.M.E., 2006b. Why are artificial polymorphous concepts so hard for birds to learn? Quart. J. Exp. Psychol. 59, 251–267.
- Loidolt, M., Troje, N.F., Huber, L., Aust, U., Fieder, M., 1997. The role of texture and shape in complex visual perception in pigeons. In: Taborsky, M., Taborsky, B. (Eds.), Advances in Ethology: Vol. 32. Contributions to the XXV International Ethological Conference. Blackwell, Berlin, Germany, p. 92.
- Nosofsky, R.M., Palmeri, T.J., Mckinley, S.C., 1994. Rule-plus-exception model of classification learning. Psychol. Rev. 101, 53–79.
- Peissig, J.J., Young, M.E., Wasserman, E.A., Biederman, I., 2005. The role of edges in object recognition in pigeons. Perception 34, 1353–1374.

- Roberts, W.A., Mazmanian, D.S., 1988. Concept learning at different levels of abstraction by pigeons, monkeys, and people. J. Exp. Psychol. Anim. Behav. Process. 14, 247–260.
- Rumelhart, D.E., Zipser, D., 1985. Feature discovery by competitive learning. Cogn. Sci. 9, 75–112.
- Soto, F.A., Wasserman, E.A., 2010. Error-driven learning in visual categorization and object recognition: a common-elements model. Psychol. Rev. 117, 349.
- Troje, N.F., Huber, L., Loidolt, M., Aust, U., Fieder, M., 1999. Categorical learning in pigeons: the role of texture and shape in complex static stimuli. Vis. Res. 39, 353–366.
- Vaughan, W.J., Greene, S.L., 1984. Pigeon visual memory capacity. J. Exp. Anal. Behav.: Anim. Behav. Process. 10, 256–271.
- Wasserman, E.A., Kiedinger, R.E., Bhatt, R.S., 1988. Conceptual behavior in pigeons: categories, subcategories, and pseudocategories. J. Exp. Psychol. Anim. Behav. Process. 14, 235–246.
- Watanabe, S., Sakamoto, J., Wakita, M., 1995. Pigeons' discrimination of paintings by Monet and Picasso. J. Exp. Anal. Behav. 63, 165–174.

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc



Chasing sounds

Julie J. Neiworth*

Department of Psychology, Carleton College, Northfield, MN 55057, United States

ARTICLE INFO

Article history: Received 30 August 2012 Received in revised form 20 November 2012 Accepted 20 November 2012

Keywords: Monkeys Auditory Music Pattern Tamarins Language Calls

alls

1. Introduction

Recent research on auditory perception in monkeys has uncovered interesting and sophisticated cognitive processing to naturalistic stimuli and more complex melodies. In earlier work, several monkey species including capuchin (Cebus apella) and rhesus (Macaca mulatta) representing New World and Old World groupings struggled to discriminate simple auditory dimensions like absolute pitch in discrimination tasks. This outcome suggested a strong modality asymmetry between visual and auditory simple stimuli in monkeys (Cowey, 1968; D'Amato and Salmon, 1982; Wegener, 1964; Poremba et al., 2003). More recently, researchers have found pitch-selective neurons in the auditory cortex in a New World primate, the common marmoset (Callithrix jacchus). Using single-unit extracellular recordings, Bendor and Wang (2005) found a cortical region near the anterolateral border of primary auditory cortex in the marmoset containing neurons that respond significantly to pure tones. Recent human imaging studies have revealed a cortical pitch processing region anterolateral to primary auditory cortex (Penagos et al., 2004). These findings suggest similar brain processing of pitch in humans and at least some monkey species. They also suggest that monkeys may not readily demonstrate a grasp behaviorally to learn differences between very simple pure tones although their brains code for the differences. Alternatively, more complex tonal sequences which match

ABSTRACT

Prior work with Wright and others demonstrated that rhesus monkeys recognized the relative relationships of notes in common melodies. As an extension of tests of pattern similarities, tamarins were habituated to 3-sound unit patterns in an AAB or ABB form that were human phonemes, piano notes, or monkey calls. The subjects were tested with novel sounds in each category constructed either to match the prior pattern or to violate the prior habituated pattern. The monkeys attended significantly more to a violation of their habituated pattern to a new pattern when human phonemes were used, and there was a trend difference in attention toward pattern violations with melodies. Monkey call patterns generated a variety of behavioral responses, were less likely to show habituation, and did not generate a strong attention reaction to changes in the patterns. Monkeys can extract abstract rules and patterns from auditory stimuli but the stimuli, by their nature, may generate competing responses which block processing of abstract regularities.

© 2012 Elsevier B.V. All rights reserved.

better monkeys' natural calls and sounds may solicit more cognitive processing and a demonstration of their cognitive prowess with auditory stimuli.

Constructing an auditory sequence with absolute changes among the tones defines a melody. The more abstract structure of a melody is the relative relationship, the rise and fall of individual tones with respect to each other. When a melody is transposed or when its absolute frequencies are changed but its frequency relations are preserved, humans perceive the transposed melody as similar to the original one because the melody contour is identical. Perception of such relationships between frequencies, or relative pitch perception, is prominent for humans from early stages of development (Chang and Trehub, 1977; Demany and Armand, 1984; Trehub et al., 1984, 1987). But is this a primate-general ability?

Early studies of melody contours by D'Amato (1988) and D'Amato and Salmon (1982, 1984) in cebus monkeys suggested the use of absolute cues to discriminate melodies and thus a restricted use of relative relationships because the monkeys seemed to use local cues for discrimination. Others found similar use of absolute cue strategies in rhesus monkeys (Moody et al., 1986). However, Izumi (2001) trained 3 Japanese monkeys (*Macaca fuscata*) to detect changes from rising to falling contours of 3-tone sequences and the monkeys were able to transfer relative pitch perception to novel octave-shifted sequences, if only within the absolute frequency range of the training phase.

Before the Izumi study, Wright and I and several colleagues explored rhesus monkeys' understanding of relative relationships among tones (Wright et al., 2000) by constructing melodies that



^{*} Tel.: +1 507 222 4372; fax: +1 507 222 7005. *E-mail address:* jneiwort@carleton.edu

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.009

were recognized by humans to have strong gestalt or grouping properties (i.e., childhood songs like "Happy Birthday"), and by testing monkeys who had been trained to judge whether a sound was the same or different from previously heard sounds. Same-different judgments are considered more abstract in that the idea of sameness is typically based on category membership in which many items may be slightly different but they carry a common structure or organization. Same/different discriminations preclude the use of absolute unique singular features, unless one focuses on absolute differences in the decision, which is antithetical to "sameness". We found that the two rhesus monkeys in the study could readily generalize, or identify as "same" strong melodies that were transposed 1 or 2 octaves from the original presentation. This generalization failed when the melodies were constructed more randomly, and thus a strong internal structure was needed in order for the monkeys to "hear" transposed melodies by whole octaves as the same as the original melodies. This study demonstrated that monkeys can treat melodies as whole units and can generalize particular shifts in the same way that humans do. It also showed several limitations by monkeys in doing so: both based on the internal structure of the melody, and based on the type of shift that was imposed.

Since my contribution to some of this work with Wright, I have often considered how monkeys process auditory stimuli we present to them. The fact that the structure of auditory stimuli influences strongly how monkeys perceive and categorize sounds suggests that there are potentially resources used by monkeys for listening to monkey calls (their own communication system), for naturalistic sounds (like other species' communication sounds), and for music that have gone on untapped and not measured in our experiments thus far. Matthews and Snowdon (2011) found that cotton top tamarins (Saguinus oedipus), a New World monkey species that I also study, recognize the "long" calls of tamarin relatives for approximately 4 years, as evidenced by lower arousal levels to their calls than to other tamarins they may have heard during the same time period. The long call, constructed as a rising series of continuous sound, must be processed very particularly for relatives of monkeys for the habituated reaction to be maintained to individuals' calls. Snowdon also examined tamarins' reactions to cello music composed to match agitated monkey calls or relaxed monkey calls, and found the similar emotional reactions of agitation and relaxation when the two types were played (as described in Dingfelder, 2009). This finding suggests that music constructed with a structure similar to conspecific calls may be processed similarly, showing an emotional link connected to particular structures of sounds which may have evolved early in primates' history.

Another scientific event forced my hand to return to testing auditory stimuli when there was a retraction of a study originally published in 2002 by Hauser et al. in which tamarins showed an attentional reaction similar to very young infants to changes in phoneme (i.e., consonant-vowel sounds) sequences. In particular, in both infants and tamarins, when they were habituated to a particular sequence like AAB (or wi wi di), and that sequence was represented by very different local phoneme events (like ga ga ti, li li la, etc.), they noticed when the abstract pattern changed to ABB more so than when novel phonemes were presented in the same AAB pattern. This finding suggests that tamarins and humans share an ability to extract abstract patterns from human language sounds, and thus the perception and judgment of the regularity of human phoneme patterns is not something that is human-specific. Two issues require a response to this study: (1) are the findings replicable, given there is a question about missing data that led to the retraction in 2010 and (2) exactly how are the tamarins processing the phoneme sequences? Taking into consideration the past work on melodies and relative relationships, one would expect tamarins to be able to form an abstract category of an AAB or ABB

pattern, but the ability to form this may be related to the significance, the salience, or the psychological "strength" of the structure of the relative pieces. Recall that randomly constructed melodies using the same notes as melodies that humans readily remember were not generalized by rhesus monkeys in the former Wright et al. (2000) study. Music structured like monkey calls did provoke similar arousal reactions to the calls themselves. Thus one would expect differences in success and failure to generalize similar sequences and to note different sequences in melodies, monkey calls, and human language by tamarins.

I have spent a year testing habituation and dishabituation by tamarins to human phoneme sequences, melodies, and edited monkey calls. In every auditory case, structures resembling AAB and ABB were made. In every case, half of the subjects were exposed to an AAB sequence repeatedly, which changed locally within each session and across sessions in terms of cues used to make the sequence. The other half of the subject pool was exposed to ABB sequences repeatedly. Then a test of a novel sequence matching their habituated structure (another AAB when they had heard AAB's in training, for example) and a test of a novel sequence inconsistent with their habituated structure (an ABB sequence when they had heard AAB's in training, for example) were presented. Arousal and attention to the novel sequences were measured by eating behavior. Past research across a variety of mammalian species has indicated that food eating behavior can be altered by novel or stressful stimulus presentations (for a review, see McSweeney and Swindell, 1999), and human infants will decrease high amplitude sucking behavior to novel languages but increase it to something more familiar (Moon et al., 1993, for example). The analyses presented here are based on eating behavior with short-term limited exposure to a preferred food, which increases when subjects experience lowered arousal due to habituation and decreases substantially when subjects are attending to novel items.¹ Look rate behavior was videotaped for every training and test sequence and is being coded and scrutinized for inter-rater reliability, and will be published in another article, once the coding is complete.

2. Methods

2.1. Subjects

The subjects were 16 adult-aged cotton top tamarins, (Saguinus *oedipus*) housed in pairs or triplet groups in 3 different monkey colony rooms in the animal facility at Carleton College. There were 10 females and 6 males in the original phoneme condition, but in one case (a female), the test was not recorded properly, so only 15 subjects' data were analyzed. Three females died over the course of the study, leaving 13 monkeys in the music condition, and in the final monkey call condition. The ratio of females to males in the final two conditions was 7:6. None of the subjects had participated in an auditory discrimination experiment before this study, although they had heard humans talking and monkey calls of relatives and unfamiliar tamarins in the colony during their lifetimes. The age range of the monkeys was 5 (Egret) to 19 (Quince), with adult onset occurring around 21 months. All subjects had been adults for at least 3 years before the study began, so the age range only captures a difference in adult ages, not a developmental difference.

The monkeys were monkey-family reared in laboratory settings and had been socially housed in pairs in seven different $0.85 \text{ m} \times 1.50 \text{ m} \times 2.30 \text{ m}$ cages, with the cages visually separated by opaque sheets. The subjects were on a 12-h light-dark cycle and

¹ Look rate behavior was videotaped for every training and test sequence and is being coded and scrutinized for inter-rater reliability, and will be published in another article once the coding is complete.

had free access to water. All animals were maintained on a complete diet consisting of a yoghurt and applesauce breakfast, a lunch of Zupreem Marmoset chow, fruits and vegetables, and a protein snack (e.g., eggs, hamburger, and mealworms) daily. The protocol and care of the monkeys were approved by the Institutional Animal Care and Use Committee, and the monkeys were inspected regularly by USDA and their care met consistently animal welfare assurances.

2.2. Materials

There were 3 sets of sounds used: phoneme sequences, musical note sequences, and monkey call sequences. Some properties were the same for each sequence. Those common properties are described first, with the particular characteristic of each sound type described separately.

For each type of sequence, there were 3 sound units, separated by empty pauses. Each sound unit was edited in Audacity to last 0.5–0.7 s in duration, as was typical for spoken phonemes in the English language. Each sound unit started on a whole second, so for example, the first sound started at the 0-s mark, followed by the second sound at the 1.0-s mark, followed by the 3rd sound at the 2.0-s mark. Thus there was a 0.3-0.5 s pause between sound units within a sequence due to the length of the fully pronounced phoneme, and consequent matching musical note or monkey call. Each 3-sound sequence was played 4 times to form a trial. Within a trial each sequence presentation was started on a 5.0-s mark, so the first sequence was started at 0 s, the second sequence, at 5.0 s, and the third sequence at 10.0 s, the 4th at 15 s. Thus there was a delay of approximately 2.5 s which separated each 3-unit sequence presentation. Each trial lasted about 17-20 s in duration. All timing variations were matched exactly across the 3 sound stimuli types (phonemes, melodies, and monkey calls) such that timing itself within a sequence or across sequences was not a discriminable factor across sets. Half of the subjects heard an AAB sequence, played 4 times per trial, and half of the subjects heard an ABB sequence, played 4 times per trial. The ABB sequence used the same sound units as the AAB sequence; only the pattern was altered.

For phoneme sequences, each phoneme was constructed of a consonant–vowel combination. The phonemes used in the training set included plosive consonants (p, b, d, k, t) and nasal consonants (m, n), as well as closed front (i) and closed back (u) vowels. So for example, the AAB patterns used in training included di di bu, pu pu ki, mu mu ni, pi pi gu, bi bi du, nu nu mi, du du ki, pi pi tu, ti ti ku, and ku ku di. The ABB sequences used the same phonemes except in the ABB pattern, for example, di bu bu for the first sequence. The test phonemes were constructed with novel glide consonants (w, r, j) or liquid consonants (l), and a central open vowel (a) or a closed front vowel (i). The test phoneme sequences were la la ri and wa ji ji. All phoneme sounds were obtained from Hauser and were the female "Judy" voice used in the original 2002 study. The sequences were built using Audacity at Carleton College. They were presented using Audacity through an Apple MacBook Pro with external speakers.

The *melody sequences* were composed of 3 notes formed by the software piano included in Garage Band software. Notes were selected from frequencies from C (5th octave) (523.251HZ) to C (7th octave) (2093HZ) to match to some degree the normal pitch of monkey calls of this species, which is rather high-pitched. Within a melody sequence, each note was selected to be within an interval ranging from three to five white notes from the former note. The AAB patterns used in training included CCE, AAD, EEB, FFA, DDG, AAE, CCF, AAC, GGB, and EEB. Some of these sequences were in the 5th, 6th or 7th octave, but within that octave, the notes were within 5 white notes of each other on a piano. The ABB patterns used the same notes only matched the new pattern, so for example, the first sequence was CEE. The test sequences included GDD and FFC, both in octaves not used for those notes in previous sequences, and both descending in pattern while all training patterns were rising or ascending pitches. All sequences were formed using Audacity software, and played through Audacity on the same Apple MacBook Pro with the same external speakers.

The monkey calls were recorded in 10–15 min sessions by two undergraduate researchers during feeding times and also during a time of stress, when a confederate student donned a full face mask and marched around the colony rooms to solicit barks and warning calls. The calls were then downloaded to Audacity to be edited into 0.5–0.7 s sound units for use in sequences to match the length of the phonemes and musical notes. Attempts were made to parse at natural communicative sounds. The researchers could readily identify different "long" calls, barks, chirps, and screeches used often to indicate searches, warning calls, and food-related calls. The monkey calls that were edited included 4 different "long" calls that consisted of ascending continuous calls, 4 different monkey barks which are shorter bursts, 2 screeches evoked just before scattering following an intruder, and several chirping calls solicited by feeding times, each with multiple chirping sounds at different pitches. These were coded and mixed such that AAB training sequences included Bark1-Bark1-Chirp1, or Bark2-Bark2-Screech1, for example. ABB sequences used the same sound units but matched the appropriate sequencing, for example, the first sequence was Bark1-Chirp1-Chirp1. The test call sequences were composed of a High-Low-High (HLH) call, a double chirp (DC1), a long call (LC1) and a doublebark (DB1) which had not been used during training. They were HLH-DB1-DB1 and LC1-LC1-DC1.

2.3. Procedure

There were 3 sound conditions consisting of habituation (5 sessions) and a test (2 trials) conducted in the same series for all subjects: Phoneme habituation and testing, Melody habituation and testing, Monkey calls habituation and testing. For each condition, subjects were presented 2 trials of auditory sequences in each session for 5 consecutive sessions. Each trial consisted of 4 presentations of the same auditory sequence. Before each trial began, undergraduate researchers would position a single digital video camera on a tripod in front of each cage in which a pair or triplet of monkeys was housed. The primary investigator (PI) moved the laptop computer and speakers on a small cart behind the cages in the room to a corner of the room. On her cue, the undergraduate researchers would simultaneously enter the cages and deposit 10 Frosted Cheerios in each food bowl and exit the cage. They would immediately stand behind their cameras and begin recording. The PI would then play the first trial, which consisted of 4 presentations of the same ABB or AAB sequence, depending upon the assignment for habituation for the monkeys in the room. Each trial lasted about 17-20 s, at which point the undergraduate researchers would immediately re-enter the cages and report how many Frosted Cheerios remained. Monkeys could eat Frosted Cheerios while listening to the stimuli, but if they were surprised or were attending closely to the stimuli, their orienting behavior typically prevented them from engaging in eating. Thus if a monkey was really interested in attending to the stimulus, he may have 10 Cheerios remaining and may not have eaten any during the trial. The researchers would then repeat this process for a second trial to complete the session. When recording, the undergraduate researchers would focus on the body and head posture. Because rewards were being offered, the monkeys usually traveled to the food shelf and so were together during the recording period. The data reported here are the rates of eating Cheerios during the sound sequence presentations, recorded as a percentage of the total eaten. If no Cheerios were eaten of the 10 offered, a 0% eaten score would be entered. If 3 Cheerios of the 10 were eaten, a 30% eaten score would be entered.

Each pair of subjects was exposed to the training pattern (either ABB or AAB) for 5 sessions, with 2 trials presented per session. In the 5th session, 2 training trials were presented, followed by 2 test trials. The test trials were composed of one trial in which the sounds were novel and the pattern was novel (i.e., ABB if the subjects had habituated to AAB), and one trial in which the sounds were novel but the pattern was the same one from habituation (AAB if the subjects had habituated to AAB). The same data (look rates as indicated from video of body and head posture, and amount of Cheerios eaten) were recorded in the test trials.

For each condition, half of the subjects were habituated to AAB and the other half, to ABB. In each room, all subjects housed in that room were habituated to one type of sequence. Across rooms the sequence used for habituation varied. The phoneme habituation and test occurred in December 2011. The melody habituation and test occurred in March of 2012. The monkey call habituation and test occurred in August 2012. At least 3 months of time separated each habituation and test.

3. Results

For each condition (phoneme, melody, and monkey call) habituation was graphed as a function of average percentage of food eaten across the 5 habituation sessions. Fig. 1 shows the rate of habituation to phoneme sequences, to melody sequences and monkey calls, respectively.

In each condition, there was a change to more eating in the last session as compared to the first. For phoneme habituation, a best-fitting linear trend indicates a 1% increase per session, with the eating rate starting at 32.4% ($R^2 = 0.77$). In contrast, for melody habituation, the eating rate was suppressed in Session 1 (10.7%). Anecdotally, we observed that the monkeys seemed more agitated and attentive initially to the melody sequences. A best-fitting linear trend indicates a 5% increase in eating per session ($R^2 = 0.72$). The monkey call sequence was the only auditory set that generated a sensitization effect because Session 2 generated a much lower rate of eating (12.86%) than Session 1 (22.86%). Overall, the best-fitting linear trend for data from monkey calls shows a 3% increase in eating across sessions, but with only 38% of the variance accounted for by the line ($R^2 = 0.38$). The rates of habituation to the same number of exposures of the same duration sequences varied depending upon the nature of the sound. Phoneme sequences and melody sequences produced more consistent slow habituation for which increasing functions were noted, and most of the variance was accounted for by best-fitting lines. The monkey call sequences produced the most varied arousal, with a fairly poor fit from a linear trend analysis.

A separate repeated measures analysis of variance (ANOVA) was conducted using average percent food eaten per subject as the



Fig. 1. Average percent Cheerios eaten during the phoneme sequence, the melody sequence, and the monkey calls across sessions.



Fig. 2. Average amount eaten to the last two training trials in habituation compared to the two novel test trials, one which also violated the pattern (different category) and one which matched the habituated pattern (same category). Error bars show standard deviations.

dependent variable to 3 trial types (trained set, test with same pattern, test with different pattern) for each condition (Phoneme, Melody, and Monkey calls). Fig. 2 shows the mean and variance, represented as standard deviations, to the last 2 training trials, the test with the same pattern, and the test with the different pattern, across these 3 conditions.

The ANOVA comparing the amount eaten across the last 2 training trials and the two test trials in the Phoneme condition was significant, F(2, 14) = 4.04, p = 0.04. By pairwise comparisons between the trained set and the two tests, there was significantly less eaten when the test violated the prior habituated pattern (mean = 24.58%) as compared to the amount eaten to the trained set (mean = 36.78%), p = 0.02. The difference in amount eaten between the trained set and the test which maintained the same pattern was not significant, p = 0.12. Thus the monkeys hesitated more and subsequently ate less when presented a novel sequence that also violated the pattern to which they had habituated when the stimuli were human phonemes.

The ANOVA comparing the amount eaten across training and the two tests in the melody condition was not significant (p = 0.28). By pairwise comparisons between the trained set and the two tests, there was a trend difference between eating rates to the novel test which violated the prior pattern (mean = 24.29% eaten, p = 0.077) and the trained set (mean = 32.5%). All other comparisons were not significant. The ANOVA testing the amount eaten against the trained and test conditions in the monkey call condition was also not significant (p = 0.41) and none of the pairwise comparisons between the trained set and the two tests was significant nor a trend.

4. Discussion

This experiment tested whether tamarins habituate to regular repeated patterns of auditory stimuli such that a violation of the former pattern produces dishabituation. It is clear from these results that tamarins can extract an abstract regular pattern, but the extent to which they can do so depends upon the nature of the auditory stimuli. With human phonemes spoken by a female voice, the tamarins showed a relaxed response from the first presentation, and only marginally habituated more to the sounds over a 5-session training period, as evidenced by a 1% average increase in eating treats per session, and a rise from 32 to 36% eating time to presentations. More importantly though, when the pattern of consonant–vowel units changed (either from AAB to ABB or from ABB to AAB), the tamarins showed a significant drop in eating during stimulus presentation. Their need to orient to the new pattern and process it as distinctly different consequently yielded less time to eat the treats.

But how should we interpret this finding? Does it suggest that tamarins show a sensitivity to human phoneme patterns shared with very young human infants? Does it suggest a more general sensitivity to regularized patterns of auditory stimuli that may be applied to any sounds and also happens to work with human phonemes? By comparing dishabituation outcomes across several auditory categories and by considering the relevance of the auditory source to the subject and the rate of habituation by the source, it should be possible to narrow the possible explanations for this phoneme pattern sensitivity effect.

First, the same level of dishabituation, as evidenced by decreased eating rates, was not found for melodies and monkey calls with the same strength as it was found for human phonemes. Thus, the phoneme pattern sensitivity that the monkeys showed is not based on a general statistical pattern recognition strategy applied to any pattern of auditory stimuli equivalently. Rather, a violation of a habituated auditory pattern generates more attention by tamarins to human phonemes than to melodies, and less to novel changes in monkey call sequences.

Why can tamarins sometimes note pattern changes in complex auditory stimuli and sometimes not? One clue may reside in the habituation rates to each class of stimuli. The phoneme sequences did not generate a strong initial attention reaction. In contrast the melody sequences suppressed eating rates to 10%, a full 20% lower than the reaction tamarins had to phonemes. It's possible that anxious and agitated responses block cognitive processing of the pattern of events more than the events themselves. This would explain more arousal to new phoneme patterns following phoneme habituation, and a more ambiguous arousal reaction to melody pattern changes because melodies, and possibly the notes themselves, originally produced more arousal.

The most variability in response happened during monkey call habituation, in which monkeys reacted with suppressed eating and increased eating in a fairly random pattern, with only slight habituation noted (3% increase per session of eating, but only 38% of the data accounted for by this linear analysis). I can report that the monkey calls had to be edited to fit the time window to match the other stimuli but strong effort was made to select full complete monkey call sound bytes to the extent that we recognize full complete calls and can code specific calls by our own hearing of them and their elicited reactions. Still, the conspecific calls we selected were mixed in ways which produced triplets of calls not normally uttered so close temporally, i.e., a food-related call coupled with territorial barks or screeches. Anecdotally, the monkeys behaved stereotypically to the calls by running in circles and scattering to screeches, and to replying with long calls to partial long calls played to them. In this case, the monkeys' cognitive processing was divided between processing what they heard and reacting to it. In a limited capacity model for attention, this should produce poor quality cognitive processing of more abstract elements, such as how the individual sounds produce a pattern.

The very different attention demands placed on the tamarins by the different categories of auditory stimuli produced the successes and failures by monkeys to notice pattern violations in this study. Human language is heard all the time and does not generate a strong alerting response, thus more subtle qualities like the patterns of sounds being perceived may be more easy to note and remember. Melodies made from piano notes are more foreign to the tamarins and, in this case, were not necessarily constructed to make a strong gestalt unit. The tamarins noted when the pattern changed, but less well than with human phonemes. The monkey calls cobbled together as ABB or AAB patterns used communicative signals in novel presentations to the monkeys, and rather than note the pattern, they reacted to the information contained in the discrete units (long call response, or territorial or hiding response). The lesson from this study, as was true of many of Wright's visual concept learning experiments, is that testing a variety of stimuli is very important, as is considering the meaningfulness or relevance of different types of stimuli to the monkey subject. A natural monkey call, if presented clearly, may induce single unit processing due to the relevance to the subject and this processing may block subtle or abstract cognitive processing, especially if the sound invokes a plan to action that divides the monkey's attention. In contrast, a limited simple auditory stimulus may not gain enough attention by the monkey to show learning or discrimination. To determine whether monkeys can notice pattern regularities or gestalt wholeness in melodies, calls, music, or language requires a consideration of the context generated by the sounds and the meaning the sounds provoke.

Acknowledgements

This research was funded in part by an NIH AREA grant # R15DH072751 to the author. The author would like to thank Anthony A. Wright for his collaborations and creativity. She would also like to recognize the following undergraduate researchers who helped with this project: Dee Dee Rupert, Rae Wood, Owen Alldritt, Deborah Tan, Jordan Palacios, Caleb Hyde, Sofia Levinson, and Evia Zack, Carleton College.

References

- Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165.
- Chang, H.W., Trehub, S.E., 1977. Auditory processing of relational information by young infants. J. Exp. Child Psychol. 24, 324–331.
- Cowey, A., 1968. Discrimination. In: Weiskrantz, L. (Ed.), The analysis of behavior change. Harper & Row, New York.
- D'Amato, M.R., 1988. A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. Music Percept. 5, 453–480.
- D'Amato, M.R., Salmon, D.P., 1982. Tune discrimination in monkeys (Cebus *paella*) and in rats. Anim. Learn. Behav. 2, 126–134.
- D'Amato, M.R., Salmon, D.B., 1984. Processing of complex auditory stimuli (tunes) by rats and monkeys (Cebus apella). Anim. Learn. Behav. 12, 184–194.
- Demany, L., Armand, F., 1984. The perceptual reality of tone chroma in early infancy. J. Acoust. Soc. Am. 76, 57–66. Dingfelder, S., 2009. Monkeys mostly like monkey music. APA Monit. 10 (November),
- 10.
- Hauser, M.D., Weiss, D., Marcus, G., 2002. Rule learning by cotton-top tamarins. Cognition 86, B15–B22.
- Izumi, A., 2001. Relative pitch perception in Japanese monkeys (Macaca fuscata). J. Comp. Psychol. 115, 127–131.
- Matthews, S., Snowdon, C.T., 2011. Long term memory for calls of relatives in cotton top tamarins. J. Comp. Psychol. 125, 366–369.
- McSweeney, F.K., Swindell, S., 1999. General-process theories of motivation revisited: The role of habituation. Psychol. Bull. 125, 437–457.
- Moody, D.B., May, B., Cole, D.M., Stebbins, W.C., 1986. The role of frequency modulation in the perception of complex stimuli by primates. Exp. Biol. 45, 219–232.
- Moon, C., Cooper, R.P., Fifer, W.P., 1993. Two-day olds prefer their native language. Infant Behav. Dev. 16, 495–500.
- Penagos, H., Melcher, J.R., Oxenham, A.J., 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 24, 6810–6815.
- Poremba, A., Saunders, R.C., Crane, A.M., Cook, M., Sokoloff, L., Mishkin, M., 2003. Functional mapping of the primate auditory system. Science 299, 568–572.
- Trehub, S.E., Bull, D., Thorpe, L.A., 1984. Infants' perception of melodies: The role of melodic contour. Child Dev. 55, 821–830.
- Trehub, S.E., Thorpe, L.A., Morrongiello, B.A., 1987. Organization processes in infants' perception of auditory patterns. Child Dev. 58, 741–749.
- Wegener, J.G., 1964. Auditory discrimination behavior of normal monkeys. J. Aud. Res. 2, 81–106.
- Wright, A.A., Rivera, J.J., Hulse, S.H., Shyan, M., Neiworth, J.J., 2000. Music perception and octave generalization in rhesus monkeys. J. Exp. Psychol. Gen. 129, 291–307.

Contents lists available at SciVerse ScienceDirect

Behavioural Processes

ELSEVIER



journal homepage: www.elsevier.com/locate/behavproc

Discrimination of coherent and incoherent motion by pigeons: An investigation using a same-different motion discrimination task^{*}

Masako Jitsumori^{a,*}, Noriyuki Nakamura^a, Edward A. Wasserman^b

^a Department of Cognitive and Information Sciences, Faculty of Letters, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan ^b Department of Psychology, Delta Center, The University of Iowa, Iowa City, IA 52242, United States

A R T I C L E I N F O

Article history: Received 18 June 2012 Received in revised form 4 October 2012 Accepted 8 October 2012

Keywords: Motion discrimination Same-different concept Gestalt Relative motion Pigeons

ABSTRACT

We trained seven pigeons to discriminate arrays of 8 identical icons that made small random movements in the *same* direction (coherent movement) from arrays of 8 identical icons that made small random movements in *different* directions (incoherent movement), with each icon moving within its own cell in an invisible 4×4 grid. During initial training, one specific configuration of icons (a fixed array) was used. The pigeons learned this discrimination and were later trained with successively introduced novel fixed arrays, and finally with novel arrays of random spatial arrangements (random arrays). Four pigeons successfully learned the final version of the task and were tested with random arrays containing different numbers of icons (from 2 to 12). Discrimination accuracy rose as the number of icons increased. These and other findings suggested that the pigeons had discriminated the visual displays by relying on the relative motion of the icons. Nevertheless, motion signals from individual icons (i.e., absolute motion) did interfere with discriminative performance to arrays of coherently moving icons. These results were considered in light of findings from another experiment in which pigeons had to search for a static icon among identical icons that moved coherently or incoherently as in the present study.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

For animals living in a world full of moving objects, motion is one of the most important visual cues for recognizing a variety of objects, such as prey, predators, rivals, and partners, most of which are in more or less constant motion. Laboratory studies have found that pigeons can: (1) discriminate the velocity of visual stimuli (Hodos et al., 1975; Mulvanny, 1978; Siegel, 1970), (2) track moving targets (Pisacreta, 1982; Rilling and LaClaire, 1989; Wilkie, 1986), and (3) discriminate complex motion such as Lissajous figures (Emmerton, 1986), rotating spiral patterns (Martinoya and Delius, 1990), and images of a clock hand moving at a constant velocity (Neiworth and Rilling, 1987).

More recently, pigeons have been found to categorize different video images of conspecifics based on their actions, such as pecking, circling, pacing, and walking (e.g., Dittrich et al., 1998; Jitsumori et al., 1999). Goto and Lea (2003) found that pigeons can

* Corresponding author.

E-mail address: mjitsu@L.chiba-u.ac.jp (M. Jitsumori).

discriminate rightward from leftward moving backgrounds, without any other spatial cues such as different patterns across time or trajectories of movement. Using video animations as stimuli, Cook et al. (2001) found that pigeons can discriminate the motion of "through-and-around" object-like stimuli (see also Cook and Roberts, 2007). Koban and Cook (2009) found that pigeons can learn to discriminate video images of objects rotating right or left around the *y*-axis, and Cook et al. (2011) found that pigeons can learn to discriminate fast- from slow-moving objects rotating around the *y*axis and later transfer the velocity discrimination to novel objects, novel rotating speeds, and novel types of motion (rotation around the *x*-and *z*-axis). These findings collectively document that pigeons can learn *specific* motion cues that may be shared by different video images.

We designed the present project to study same-different motion discrimination in pigeons. Same-different motion discrimination does not involve learning about *specific* motion cues; rather, it involves learning about the *relative* motion of two or more objects. Most studies of same-different discrimination behavior have concerned static stimuli that are simultaneously or successively presented. Wright and his colleagues (e.g., Sands and Wright, 1980; Santiago and Wright, 1984; Wright et al., 1984; see also Jitsumori et al., 1988) successfully trained animals (monkeys and pigeons) in a multiple-item memory task (a serial probe recognition task), in which a list of items is presented sequentially, and

^{*} This research was supported by Grant 13610076 from the Ministry of Education, Science, and Culture to M.J. We thank Tomoyuki Oya and Tomoya Sudo for assistance in conducting the research and Mitsuo Hanada for programming the motion stimuli. Special thanks go to Tomokazu Ushitani for his valuable comments on an earlier draft of the article.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.10.004

after short delay, a single probe item is presented that is either from the list (Same) or not from the list (Different); the subject's task is to classify the probe item as either Same or Different. Pigeons and other animals have also been successfully trained by additional researchers to discriminate simultaneous arrays of same from different visual items (e.g., Castro et al., 2010; Cook et al., 1997; Wasserman et al., 2002, 1995) or successive lists comprising same or different items (Young et al., 1999). In each case, discriminations often improve with increasing numbers of items in the training set (e.g., Katz et al., 2002; Wright et al., 1988) and with increasing numbers of items in a display (e.g., Wasserman et al., 2000; Young et al., 2007; see Cook and Wasserman, 2012; Wasserman and Young, 2010; Young et al., 1997, for reviews and interpretations of these findings).

There is currently little or only weak evidence that pigeons can discriminate same from different motion. Goto et al. (2002) trained pigeons to discriminate stimuli containing four randomly moving dots from similar stimuli in which one dot moved non-randomly toward one of the three randomly moving dots. The pigeons had difficulty detecting such patterned motion embedded within random patterns of motion. The authors deemed this motion discrimination task to be a kind of same-different concept discrimination between patterned and random motion; they argued that the number of moving dots in their study might have been inadequate to permit pigeons to discriminate between the different movement patterns.

Bischof et al. (1999) trained pigeons to discriminate large groups of dots coherently moving in a common direction (up, down, left, or right) from dots randomly moving in one of four directions. Because every dot disappeared after a brief temporal interval and reappeared at another randomly chosen location, the authors argued that efficient detection of coherent motion requires the integration of motion signals generated over a brief interval for each coherently moving dot (local motion integration) in addition to the integration of the local motion signals into a global motion percept (global motion integration). Although pigeons easily discriminated between the different motion stimuli (complete coherence vs. random motion), subsequent tests revealed that the pigeons were inferior to humans in their ability to detect coherent motion embedded in a background of random motion noise. Bischof et al. proposed that pigeons' inferior motion sensitivity may be attributed to poorer spatiotemporal motion integration at both the local and global levels (also see Nguyen et al., 2004).

In an earlier, unpublished study, we trained four pigeons to discriminate motion arrays comprising 8 identical icons randomly distributed in an invisible 4 × 4 grid on each individual trial (i.e., the arrays were created by populating a random 50% of the 16 cells). The icons made small random movements in the same direction (coherent movement) on same-motion trials, whereas the icons made small random movements in different directions (incoherent movement) on different-motion trials. The movements never took an icon out of its home cell, so there was no overlap of movement ranges or collision of icons. We hoped that the icon arrays might be suitable for pigeons to integrate motion signals, in contrast to the motion stimuli used by Bischof et al. (1999), in which coherent motion signals were generated over brief temporal intervals and at widely spaced locations in noisy random dot displays. However, all of our pigeons exhibited considerable difficulty discriminating same-motion from different-motion arrays. Even after 6 months of training, they performed particularly poorly on same-motion trials.

In the present study, we first trained experimentally naïve pigeons with a single 8-icon array, rather than with random arrays generated on each trial as in our unpublished study. The eight populated cells were matched in the coherently (same) and incoherently (different) moving arrays. From our own visual experience, the coherence of a same-motion display as a single, moving whole (or Gestalt) promotes the ready discrimination of same-motion from different-motion displays. If pigeons are able to integrate the motion of the icons in much the same way as we humans do, then they should easily discriminate same-motion from differentmotion displays, relying on the particular visual pattern that is formed by the array of icons moving in unison. The design of this training procedure was based on the assumption that the same-motion icon arrays that were randomly generated on each individual trial in our previous study could have made it difficult for our pigeons to classify these "stable," but highly variable visual patterns into the same stimulus class and to discriminate them from different-motion arrays.

In order to train our pigeons to discriminate same-motion from different-motion displays, disregarding all cues tied to the specific training arrays, we successively introduced novel arrays one at a time. Finally, we trained the pigeons with arrays that were randomly generated on each trial as in our unpublished study. This regimen should have promoted the pigeons' categorizing the motion displays in terms of the presence or absence of a stable visual pattern and/or the relative movements of the individual icons. For those pigeons that successfully learned to discriminate same-motion from different-motion patterns with randomly configured arrays, we tested their transfer to new random arrays comprising fewer or more than 8 icons.

The questions that we addressed were: (1) can pigeons discriminate same-motion from different-motion arrays, despite the arrays comprising identical visual items and the arrays involving random spatial arrangements of the depicted items and (2) does same-motion versus different-motion discrimination depend on the number of visual items in the array? The results of our research will be discussed in regard to motion processing by pigeons.

2. Method

2.1. Subjects

The subjects were seven experimentally naïve pigeons kept between 80% and 85% of their free-feeding body weights throughout the experiment. Water and grit were freely available in the individual home cages.

2.2. Apparatus

The experiment used four identical operant conditioning chambers $(35 \text{ cm} \times 29 \text{ cm} \times 35 \text{ cm})$. All of the stimuli were presented on TFT color monitors (Sharp LL-T1510R) which were visible through a window $(10 \text{ cm} \times 16 \text{ cm})$ located in the middle of an aluminum panel in front of each chamber. The bottom edge of the viewing window was 20 cm above the chamber floor. The monitor was 1.5 cm behind an infrared touch frame (Carroll Touch). Between the frame and the surface of the monitor was a thin Plexiglas sheet that shielded the monitor from direct contact. The icon arrays appeared in a 200 \times 200 pixel (approximately 6.0 cm \times 6.0 cm) central square display area that could be lighted white on a black background. There were two round report areas (2 cm diameter) whose centers were 6 cm to the left and 6 cm to the right of the central square display: the left report area could be lighted green and the right report area could be lighted red. A food aperture $(7 \text{ cm} \times 6 \text{ cm})$ in the middle of the front panel afforded pigeons access to a solenoid-operated food tray containing a mixture of grains. A houselight (3W) in the center of the ceiling dimly illuminated the chamber.

The chambers and the video monitors were located in a darkened testing room. Computer programs driving the presentation of video stimuli and controlling the houselight and feeder were developed in Direct X 7.0 and Visual Basic 6.0 (Microsoft).



Fig. 1. Examples of initial directions of movement of icons in same-motion and different-motion displays.

2.3. Stimuli

A "checked box" icon $(14 \times 17 \text{ pixels})$ was chosen from which to construct arrays of 8 icons; the icons were identical for both samemotion and different-motion arrays. The 8 icons in an array were located in eight predetermined areas within an invisible 4×4 grid, with each icon in the center of its grid area. The grid areas were then pseudo-randomly moved left, right, up, or down by 0 to 25 pixels, with the constraint that, in the resulting array, the icons did not lie in orderly rows or columns and the grid areas in which the icons could move $(47 \times 47 \text{ pixels})$ did not overlap one another.

Fig. 1 depicts the first array (Array A) that we showed to the pigeons, with the arrows indicating examples of the initial directions of movement on same-motion (left panel) and differentmotion (right panel) trials. Each icon began at the center of its movement area from which it moved at a rate of 37.5 pixels per s. On same-motion trials, each of the icons moved in the same direction, selected randomly from 16 possible directions (range = 0 to 360° , each directional step = 22.5°). On different-motion trials, each icon's movement direction was randomly determined without repetition. After the icon had moved 20 pixels, a new direction was randomly selected from the possible 16 directions (i.e., a single direction was selected for all the icons on same-motion trials and eight different directions was selected for the icons on differentmotion trials), with this 20-pixel selection process repeated to the end of the stimulus presentation. If an icon "bumped" against an edge of its grid area, then its path changed in the mirror-reflected direction. Thus, each icon moved about in its grid area in a highly erratic fashion on both same-motion and different-motion trials.

We prepared six different 8-icon configurations (Arrays A, B, C, D, E, and F) which are shown in Fig. 2. These arrays were successively introduced during the early training phase and will be referred to as *fixed* arrays. During the final training phase, we generated new stimulus arrays on each individual trial, with the 8 icons randomly distributed to 8 of the 16 possible areas. These arrays will be referred to as *random* arrays. For testing arrays involving fewer or more than 8 icons, we also *randomly* distributed the icons on each individual trial. For any given same-motion array, the icons moved incoherently. Readers can view examples of the same-motion and different-motion displays at: http://cogsci.L.chiba-u.ac.jp/pigeon/SDmotion/profile.html.

2.4. Procedure

2.4.1. Initial training

We first trained the pigeons to peck at a black cross in the center of a white display area ($6 \text{ cm} \times 6 \text{ cm}$), using conventional handshaping. We then trained the birds to peck the red and green report areas. Once we established responses to all of these areas, we added the same-motion and different-motion displays derived from the



Fig. 2. Fixed arrays of 8 identical icons sequentially introduced during training.

fixed array (Array A in Fig. 2) that was the template for creating the moving discriminative stimuli. A peck to the black cross turned on a stimulus array as a black-on-white movie. A single peck anywhere in the display area illuminated only one report area—the green (left) or the red (right) report area—that was to be correct in the following discrimination training phases (the assignment of the report areas varied across the pigeons). A response to the report area black-ened the display and the report areas and immediately delivered food (the food hopper was presented for 3 s). Intertrial intervals (ITIs) averaged 8 s (range: 6-10 s), during which the houselight was illuminated. This training phase lasted for 2 sessions of 48 trials each.

2.4.2. Discrimination training

Discrimination training began by simultaneously illuminating *both* correct and incorrect report areas following presentation of a stimulus array. We increased the number of pecks required to the stimulus array (observing responses) to 30. An incorrect choice response produced a 3 s blackout. Following an incorrect response, the same trial was repeated until the pigeon responded correctly. Correction trials were not scored for data analysis. Each daily session consisted of 30 same-motion trials and 30 different-motion trials (60 total trials), pseudo-randomly ordered, with the constraint that no more than three same- or three different-motion

trials could occur in succession. Other procedural details were the same as in initial training.

We first trained the pigeons with a fixed 8-icon array (Array A). So, the trials differed only in the direction of the 20-pixel movements (i.e., the starting point was consistent across trials) and whether the 8 icons moved coherently (same) or incoherently (different). When a pigeon reached 80% correct responses on both same- and different-motion trials for 2 consecutive days, a new fixed array (B) was introduced on 20 randomly selected trials (10 same- and 10 different-motion trials); on the remaining 40 trials, the old array (A) appeared equally often on same- and differentmotion trials. Training was continued until a pigeon reached 80% or better accuracy with the new array on both same- and differentmotion trials for 2 consecutive days, and performance to the old array was better than 80% accuracy on both same and different baseline trials. Another new array (C) was then similarly introduced on 20 trials; on the remaining 40 trials, the old arrays (A and B) appeared equally often on same- and different-motion trials. In this way, new arrays were introduced one-by-one until a pigeon performed at an average of 80% correct or better with the new array on the first day and also the discrimination criterion (80% or better accuracy with both the new array and the old arrays on both sameand different-motion trials for 2 consecutive days) was attained within 3 days. If a pigeon failed to reach the final stepwise criterion, then training with the given arrays was continued. When the pigeon reached the discrimination criterion, a new array was similarly introduced; that is, the pigeons received only 2 or 3 sessions with the latest novel array before reaching the final stepwise criterion

The pigeons were then trained with randomly generated samemotion and different-motion arrays. When the pigeons reached 80% correct or better responses on both same- and differentmotion trials for 2 consecutive days, the number of training trials was increased to 106 (53 same- and 53 different-motion trials). Now, correct responses on 23 same-motion trials and 23 differentmotion trials involved the food hopper being raised for only 1 s, in order to prevent the birds from overeating; correct responses on the remaining 60 trials were followed by 3-s presentations of the food hopper. Other procedural details, including the correction method, were the same as in the previous training phase with fixed arrays. After the pigeons again reached the discrimination criterion, they proceeded to testing.

2.4.3. Generalization testing

We administered 8 testing sessions. Testing sessions began with 30 assessment trials involving 15 same- and 15 different-motion arrays (for correct choices on 8 of the same- and different-motion trials, the food hopper was presented for only 1 s; on the remaining 7 trials, the food hopper was presented for 3 s). If the pigeon performed below an average of 80% correct across both same- and different-motion trials during the assessment period, then the session was recorded as a training session. If the pigeon performed at or above 80% correct in the assessment period, then testing was conducted on the following 76 trials, which were composed of two randomized blocks of 38 trials containing 14 same-motion arrays, 14 different-motion arrays, and 1 each of the 10 different types of testing arrays (same-motion and different-motion arrays each composed of 2, 4, 6, 8, and 12 icons). Therefore, standard training days sometimes intervened between testing days. Differential food reinforcement for left and right report responses was given on training trials (for correct choices on one half of the same- and different-motion trials, the food hopper was presented for only 1s; on the other half of the trials, the food hopper was presented for 3 s). Nondifferential reinforcement for left and right report responses was given on testing trials; food (3 s presentation of the food hopper) was given regardless of the pigeons' choice responses. Correction trials were implemented on training trials in the generalization test sessions including the assessment period in each session. No correction trials were necessary on testing trials.

3. Results

3.1. Discrimination training

We first trained the pigeons to discriminate same-motion displays from different-motion displays of Fixed Array A. Four pigeons (Birds 1, 3, 5, and 6) learned the discrimination relatively quickly; they took 30, 34, 20, and 18 days, respectively, to reach the performance criterion (i.e., 80% or better accuracy on both same- and different-motion trials for 2 consecutive days). The remaining three pigeons reached the criterion (Birds 2, 4, and 7), but learning was slow; they required 89, 119, and 91 days, respectively. We henceforth refer to the former pigeons as "quick learners" and the latter ones as "slow learners."

When Array B was introduced, the pigeons generally performed accurately to the new different-motion array (mean=77% correct; range = 40% to 100%), whereas the pigeons' performance to the new same-motion array precipitously fell (mean = 51% correct; range = 20% to 60% correct). For the slow learners, the second phase of training failed to improve their performance to Fixed Array B; rather, this training appeared to strengthen these birds' tendency to respond "different" to same-motion displays of Fixed Array B. These pigeons had to be dropped from the experiment after 70 additional training sessions. In the last training session, accuracy scores with same-motion and different-motion displays of Fixed Array B were 10% and 100% correct for Bird 2, 10% and 50% correct for Bird 4, and 50% and 70% correct for Bird 7, respectively. It was as though these birds were discriminating same-motion displays of Array A from all of the remaining motion displays (i.e., different-motion displays of Arrays A and B and same-motion displays of Array B). This finding suggests that the slow learners may have made the initial discrimination of Fixed Array A on the basis of the unique configuration of the coherently moving icons, which were not presented in Fixed Array B. The discrimination broke down with new coherent movement patterns characteristic of the new array, in which case the pigeons were strongly inclined to make a "different" report response.

The quick learners learned to discriminate same- and differentmotion displays of successively added novel arrays. Table 1 summarizes the transfer to novel arrays for each pigeon, including the slow learners, detailing the percentage of correct choices made on same-motion trials and on different-motion trials in the session on which each array was first presented, together with the number of required sessions (2 consecutive sessions at least) for each array. Bird 6 reached the final stepwise criterion when Array D was introduced. Bird 3 continued to perform poorly on samemotion trials, but it eventually attained the final stepwise criterion when Array F was introduced (the first array A was dropped from the pool of training arrays, and instead, F was introduced, with B, C, D, and E appearing equally often on 40 trials and F appearing on the remaining 20 trials). Bird 5 failed to reach the final stepwise criterion with Array F (it required 5 sessions to reach the criterion with Array F), but it was advanced to training with the random arrays, without training using additional new fixed arrays. Bird 1 performed accurately from the first session with Array C; its performance over the second and third sessions averaged 85% correct on both same- and different-motion trials, and this bird proceeded directly to training with the random arrays.

The two right-hand columns of Table 1 depict transfer to the random arrays. Although Bird 1, which had received discrimination training with only three fixed arrays (A, B, and C), showed an asymmetrical transfer effect (performance accuracy was far above

Table 1

Percentage of correct choices for same-motion trials and for different-motion trials in the session where each novel array was initially added into baseline training trials. Numbers in the parenthesis are the numbers of sessions required to reach performance criterion. Birds 2, 4, and 7 were dropped from the experiment after 70 training sessions with Array B.

Subject	Array B		Array C		Array D		Array E		Array F		Random	
	Same	Diff	Same	Diff								
Bird 1	40	90 (79)	70	90(3)	-	-	-	-	-	-	30	90
Bird 3	60	80 (21)	60	90(11)	30	90 (20)	70	100 (9)	80	100(2)	87	83
Bird 5	60	60 (32)	100	100(6)	90	90(5)	90	80(31)	70	100 (5)	83	93
Bird 6	60	90(9)	50	90 (32)	100	100(2)	-		-		87	87
Bird 2	60	80	-	-	-	-	-	-	-	-	-	-
Bird 4	20	40	-	-	-	-	-	-	-	-	-	-
Bird 7	60	100	-	-	-	-	-	-	-	-	-	-

Note: Same, same-motion trials; Diff, different-motion trials.

chance on different-motion trials, whereas it was below chance on same-motion trials), the remaining three pigeons showed high levels of discrimination transfer to all of the random arrays. The mean accuracy scores averaged across these three pigeons were 86% correct on same-motion trials and 88% correct on differentmotion trials. Later, the pigeons, including Bird 1, were successfully trained to discriminate same- and different-motion displays of random arrays. Birds 1, 3, 5, and 6 took 30 days, 10 days, 2 days, and 2 days, respectively, to reach the criterion of 80% correct on both same- and different-motion trials for 2 consecutive days. Then, the number of trials was increased to 106 per session.

3.2. Generalization testing

Discrimination of same- from different-motion displays with 8 icons during the 4 days before testing averaged 87% correct for the four pigeons (Birds 1, 3, 5, and 6). During the 8-day testing period, discrimination for 8-icon motion arrays on testing trials averaged a highly similar 86% correct; discrimination for 2-, 4-, 6-, and 12-icon testing arrays averaged 56%, 62%, 80%, and 96% correct, respectively. The results for each individual bird are shown in the top-left panel of Fig. 3. The four pigeons were similar in their choice accuracy, which increased as a function of the number of icons, with highest accuracy to the 12-icon arrays. The accuracy data were subjected to a trend analysis. An alpha level of .05 was adopted in this and all following analyses. It revealed that only the linear contrast was significant [F(1, 12) = 159.38, p < .001], confirming that the pigeons' choice accuracy rose as the number of icons was increased.

The top-right panel of Fig. 3 separately depicts the percentage of correct same and different responses averaged across the four pigeons, as a function of the number of icons in the displays. Choice accuracy generally rose as a function of the number of icons on both same- and different-motion trials. With the 4-icon, 6-icon, and 8-icon motion arrays, however, performance accuracy was higher on different-motion trials than on same-motion trials. Mean correct same and different responses were 53% and 70% with the 4-icon arrays, 77% and 83% with the 6-icon arrays, and 78% and 94% with the 8-icon arrays. This asymmetrical effect was not observed with the 2-icon and 12-icon motion arrays.

A repeated-measures, two-way analysis of variance was conducted, with number of icons (2 vs. 4 vs. 6 vs. 8 vs. 12) and motion (same vs. different) as independent variables. The effect of icon number was significant [F(4, 12) = 40.97, p < .001]. The effect of motion [F(1,3) = 1.08, p = .38] and the number by motion interaction [F(4, 12) = 0.95, p = .47] were not significant. Paired *t*-tests (Fisher's LSD) disclosed that 6-, 8-, and 12-icon arrays produced significantly different transfer performance from 2-icon arrays [ts(12) > 6.37, p < .001] and from 4-icon arrays [ts(12) > 4.88, p < .001]. Transfer performance also differed significantly between 8-icon and 12icon arrays [t(12) = 2.78, p < .05]. Transfer performance did not differ significantly between 2-icon and 4-icon arrays or between 6-icon and 8-icon arrays.

The bottom panels of Fig. 3 show the corresponding results from the four individual pigeons. Close inspection of the individual bird data revealed that, with 2-icon, 4-icon, 6-icon, and 8-icon displays, the pigeons generally performed more accurately on differentmotion trials than on same-motion trials, but with some exceptions for Birds 1 and 6. A peculiar finding was that Bird 6 showed a strong tendency to respond "same"; correct "same" responding (94%) was far above chance, whereas correct "different" responding (25%) was far below chance (see below for a possible explanation). When the number of icons increased from 8 to 12, the four pigeons consistently showed a facilitation effect on same-motion trials; mean accuracy scores were 78% correct and 98% correct with the 8-icon and 12-icon arrays, respectively. In contrast, no systematic change was observed on different-motion trials.

4. Discussion

The present study explored the possibility of pigeons learning a same-different discrimination involving several visually identical stimuli, each placed in motion, where the individual items in an array either moved in unison or moved independently of one another. The questions that we addressed were: (1) can pigeons discriminate same-motion arrays from different-motion arrays of identical visual stimuli? and (2) does such same-motion versus different-motion discrimination depend on the number of visual items in the arrays? Our results provide clear and affirmative answers to these questions.

As discussed in our introduction, we expected that if a samemotion display is perceived by pigeons as a single, moving whole (or Gestalt), then our pigeons should readily discriminate samemotion from different-motion displays of a single fixed array, by relying on the particular visual pattern that is formed by the array of icons moving in unison. All of the pigeons learned to discriminate the same- and different-motion displays of the first array (Array A), but 3 of the 7 pigeons (the slow learners) required large numbers of sessions to do so.

Performance to a novel fixed array (Array B) indicated that the slow learners had discriminated the motion displays of Array A by relying on the unique spatial configuration of the coherently moving icons. It seems that processing the spatial configuration of the icons required considerable cognitive resources for the slow learners, a possibility which does not favor of the idea that early perceptual mechanisms for integrating the motion of coherently moving items (Gestalt) enable pigeons to readily perceive a visual pattern. Yet, one might propose that some behavioral factors, such as observing distance (i.e., visual angle), might have strongly interfered with these birds perceiving a same-motion display as a single, moving whole.



Fig. 3. Top-left panel: Percentage of correct choices as a function of the number of icons in the motion array for 4 individual pigeons. Top-right panel: Mean percentage of correct *same* and *different* responses as a function of the number of icons in the motion array across the 4 pigeons. Error bars indicate standard error of the mean (SEM). Bottom four panels: Percentage of correct *same* and *different* responses as a function of the number of icons in the number of icons in the motion array for 4 individual pigeons.

The remaining four pigeons (the quick learners) may also have used the specific configuration of coherently moving icons to discriminate the motion displays early in training (Table 1), but they eventually discriminated the same-motion arrays from the different-motion arrays, despite the arrays involving random spatial arrangements of identical visual icons. This finding suggests that the quick learners may have used a generalized rule instead of a particular spatial arrangement of icons to solve the same-different motion discrimination.

There are at least four theoretical explanations for the discrimination behavior of the quick learners. First, the quick learners may have responded discriminatively to the same-motion and differentmotion displays on the basis of the variability in movement of the individual icons in the array-motion entropy discrimination. Second, the quick learners may have classified the stimuli into stable visual patterns (same-motion arrays) and unstable visual patterns (different-motion arrays)-Gestalt formation. Third, the quick learners may have come to discriminate the motion displays based on the same or different movements of the icons in a display-a generalized same-different concept. Finally, the quick learners may have discriminated the motion displays by relying on the perceived relative motion of two or more icons in a display-perception of relative motion. Any of these explanations of the pigeons' samedifferent motion discrimination behavior would have to account for improved performance with increasing numbers of icons, including the significant improvement in performance shown by the pigeons when the number of icons increased from 8 to 12, as discussed ahead.

Motion entropy. With static visual stimuli, it has repeatedly been found that the amount of variability or entropy in multi-element arrays plays a critical role in pigeons' discriminating same from different arrays; this conclusion holds when visual arrays comprise different *numbers* of items as well as when arrays comprise different *mixtures* of same and different items (Wasserman and Young, 2010; Wasserman et al., 2000; Young et al., 1997, 2007).

Following this logic, our pigeons might have responded discriminatively to the same-motion and different-motion displays on the basis of the variability in the movement of the icons in the arrays. If this were so, then we would predict that performance accuracy should improve as the number of icons is increased on differentmotion trials (because motion entropy rises as the number of icons increases), but not on same-motion trials (because motion entropy always equals 0, regardless of the number of icons). Our results with dynamic stimuli do not support this notion because accuracy increased on both same-motion trials and different-motion trials (also see Castro et al., 2010, in press; Castro and Wasserman, 2011 for more evidence on this matter).

Gestalt formation. We might assume that the pigeons should make a "same" report response because same-motion displays contain a stable visual pattern that does not occur on different-motion displays, in which case the pigeons should make a "different" report response. If this Gestalt notion were true, then pigeons' performance with different-motion displays should have remained high when the number of icons decreased from 8 to 6 to 4 to 2, in which case the incoherently moving icons should not have produced any stable visual patterns; however, pigeons' performance fell when the number of icons decreased from 8 to 6 to 4 to 2. Another finding that is inconsistent with the Gestalt notion is that all of the pigeons performed more accurately with 12-icon arrays than with 8-icon arrays on same-motion trials. It is unlikely that 12 coherently moving icons would produce more stable visual patterns than 8 coherently moving icons; the number of icons may well differ, but the visual patterns in these same-motion displays should nonetheless remain stable.

But, perhaps these arguments against the Gestalt interpretation are too simple. One might propose that the performance decline produced by smaller numbers of incoherently moving icons on different-motion trials is due to the generalization decrement that eventuates with any novel random array comprising a discrepant number of icons. One might further propose that larger numbers of coherently moving icons may indeed produce denser and more perceptually stable visual patterns, which might enhance the pigeons' performance on same-motion trials.

Generalized same-different concept. This interpretation assumes that our pigeons simultaneously compared the movements of at least a few individual icons and responded to the motion stimulus based on same-different judgments (serial comparisons may not be available, because each icon changed its direction after it moved 20 pixels taking only about 0.53 s). To the extent that such samedifferent conceptualization takes place, the pigeons should have successfully discriminated the same- from the different-motion displays even when the number of icons decreased to 2. Apparently, this was not the case because discriminative performance was poorest with just 2 icons in the arrays.

In the present experimental setting, however, the pigeons were very close to the displays, so that simultaneous comparisons of movement may become increasingly difficult when the distances between the icons increased with decreasing numbers of icons in the displays. If pigeons have an "attentional spotlight" (Wasserman et al., 2000), then more items should fall within its area of activation as the number of presented items increases. This idea hypothesizes that more items afford better discrimination performance because of the greater number of possible movement comparisons, even though the actual number of attended items may be (and probably is) less than the presented number (Young et al., 1997). Discrimination based on a generalized same-different concept thus expects improved performance with increasing numbers of items on both same-motion trials and different-motion trials, just as we found.

Nevertheless, this account does not fully explain our findings. Although the effect of motion (same vs. different) was not statistically significant, the four pigeons exhibited higher accuracies on different-motion trials than on same-motion trials with arrays containing 2, 4, 6, and 8 icons, except 4-icon and 6-icon arrays for Bird 1 and 2-icon arrays for Bird 6. We cannot yet specify which factors might be responsible for the asymmetry between same and different judgments with arrays containing 8 or fewer icons.

Perception of relative motion. Pigeons' early visual processing may enable them to directly perceive the relative motion of 2 or more icons that fall within the area of an "attentional spotlight." This idea assumes that pigeons discriminate same-motion from different-motion displays by relying on the impression of relative motion itself, rather than on a generalized same-different concept, which assumes that the movements of individual icons are encoded independently of one another and then compared to arrive at samedifferent judgments.

This direct impression of relative motion is presumably produced by changes in the perceived distance between the presented icons. The between-icon distance continuously changes in the different-motion displays, in which case pigeons may directly perceive the motion impression; however, in the same-motion displays, the moving icons retain the same distance from each other, in which case pigeons may not directly perceive this motion impression. The relative motion account assumes that the movements of 2 (or a few) icons are encoded into a single motion signal, in contrast to the same-different concept account which assumes that the movements of individual icons are encoded independently of one another.

The quick learners might have come to make a "different" report response when they had the motion impression and to make a "same" report response when they did not. However, each individual icon in a display does in fact move. The motion signals of the individual icons are likely to induce the pigeons to make a "different" report response (i.e., a response associated with the motion impression), which may encourage erroneous responses on same-motion trials (as discussed later, this possibility is suggested by another investigation in which pigeons were tested in a visual search task). In contrast, the motion signals of individual icons and the relative motion impression may not have such a conflicting effect on different-motion trials. Extensive training may have enhanced the quick learners to ignore the motion signals of the individual icons (i.e., the absolute motion of the individual icons). However, the perception of absolute motion, occurring in early visual processing, could have a somewhat persistent effect of interfering with performance to the arrays of coherently moving icons during generalization testing. The pigeons might thus have exhibited generally lower accuracies on same-motion trials than on different-motion trials, although not significantly so over the numbers of icons tested

When the number of icons was decreased, the discrimination relying on perceived relative motion may be more difficult on both same-motion and different-motion trials due to the increased distances between the icons. As noted earlier, Bird 6 was strongly inclined to make a "same" report response with 2-icon arrays. Perhaps the activation area of this bird's "attentional spotlight" was relatively small, so that the 2 icons were unlikely to simultaneously fall within the area (i.e., only the motion signal of a single item may exist), in which case this bird made a "same" report response.

We recently conducted a visual search study in which pigeons had to search for and peck 1 static icon (target) among 5 moving icons (distractors) in displays comprising a total of 6 identical icons (Nakamura et al., in preparation). The target did not move, but the distractors either coherently or incoherently moved in their home cells as was the case in the present study.

Pigeons successfully learned to search for and peck the target. When they were later tested with display sizes of 2, 4, 6, 8, and 10, search accuracy decreased and reaction time for correct search responses increased as a function of display size (i.e., serial search), with both coherent and incoherent motion of the distractors. In other words, the static target failed to guide attention even when a large numbers of distractors moved in unison.

In contrast, human participants, who were tested in the same experimental situation, efficiently searched for the static target among coherently moving distractors (i.e., parallel search, with the slope of the reaction time × display size function near 0), while their reaction time scores increased as a function of the number of incoherently moving distractors (i.e., serial search). These findings in humans are consistent with the results of previous studies showing that humans take advantage of the perceptual organization of coherently moving distractors (Royden et al., 2001; also see Ansorge et al., 2006; Driver et al., 1992; Kingstone and Bischof, 1999).

As well, the pigeons' search slope was even steeper (i.e., search was less efficient) with coherently than with incoherently moving distractors, a finding in the opposite direction from that of humans. This disparity means that when pigeons view the icons in a display one after another, the time taken to avoid an attended distractor is longer when it moves coherently than when it moves incoherently with the others.

To examine the effect of the distances among the distractors on the search for the static icon among coherently or incoherently moving distractors, search accuracy in the last training sessions (display size = 6) was analyzed in terms of the mean distance among the distractors (the distance between each distractor pair was measured by the center-to-center distance of their home cells). Search accuracy increased as the mean distance among the distractors in a display decreased, but this performance improvement was larger with the incoherently than with the coherently moving distractors.

This finding suggests that, although the pigeons used the motion signal of the attended distractor to avoid pecking it during the search for a static icon, they also made use of the perceived relative motion of the attended distractor and other distractors in its neighborhood. When the attended distractor moved coherently with other distractors, the pigeons were inclined to incorrectly respond to it during the search for the static icon. The less efficient search with coherently moving distractors was due to the pigeons' difficulty detecting the motion of the attended distractor when it moved coherently with others in its neighborhood.

If pigeons' early perceptual mechanisms enable them to see a same-motion array as a single, moving Gestalt, then the pigeons might efficiently search for the static target among coherently moving distractors when the number of distractor is sufficiently large to produce a perceptually stable visual pattern. Apparently, this was not the case; the coherently moving distractors did not promote an efficient search, but instead interfered with the search for the target compared to the same number of incoherently moving distractors, with the discrepancy in reaction time increasing as the number of distractors increased. These findings from the visual search investigation thus suggest that *Gestalt formation* may not be a plausible account for the pigeons' same-different motion discrimination in the present study.

Our conclusion is that the *perception of relative motion* is the most plausible and parsimonious account for the findings in generalization testing. That is, the impression of the relative motion of at least a few icons enabled the quick learners to discriminate the same-motion and different-motion displays involving random spatial configurations of icons as well as to subsequently discriminate the motion displays comprising novel numbers of icons.

The quick learners also made use of the specific configuration of coherently moving icons to discriminate the motion displays in the early stages of training (Table 1). The relative motion of a few icons and the spatial configuration of the icons in an array may not represent visual cues that conflict with one another, but that jointly facilitate the discrimination of same-motion from different-motion displays. Given that pigeons may not possess the early perceptual mechanisms that are necessary for seeing a visual array as a single, moving Gestalt, processing the spatial configuration of the icons may take substantial cognitive resources. It is plausible that training with successively introduced fixed arrays eventually led these birds to discard the strategy of discriminating the motion displays by relying on the already learned spatial configurations of the icons, so that they came to discriminate the same-motion and differentmotion displays primarily on the basis of perception of relative motion.

On the other hand, the slow learners might have discriminated same-motion from different-motion displays involving the initial arrays by relying solely on the unique spatial configuration of the icons. This strategy required large numbers of sessions (99.7 session on average) to learn the discrimination with the first array, compared with the quick learners (25.5 sessions on average). With the data at hand, we cannot pinpoint the specific factors (e.g., observing distance) that caused the slow learners to be so insensitive to the relative motion of the icons.

In conclusion, the results of the present experiment clearly disclose that pigeons can accurately discriminate same-motion from different-motion displays of identical items. Pigeons can do so, perhaps without processing multiple items moving in unison as an integral whole—a phenomenon that is well documented for human observers and that was first dubbed "common fate" by Gestalt psychologists (e.g., Wertheimer, 1923).

In the present experimental setting, however, the pigeons were very close to the displays, which may be responsible for our birds' difficulty in grouping the coherent motion of separate items. Nonetheless, in the natural environment, this visually sophisticated species may see objects in the far distance, in which case coherent motion may help them to integrate separate objects into a group. Motion discrimination should be of obvious ecological relevance to avian species, particularly those that fly and cope with motion information while taking off, landing, or looking for conspecifics while navigating in flocks. For that reason and because of the paucity of information on dynamic visual perception in avian species, further research is needed to elucidate the mechanisms of motion discrimination by pigeons.

References

- Ansorge, U., Scharlau, I., Labudda, K., 2006. Visual search for a motion singleton among coherently moving distractors. Psychol. Res. 70, 103–116.
- Bischof, W.F., Reid, S.L., Wylie, D.R.W., Spetch, M.L., 1999. Perception of coherent motion in random dot displays by pigeons and humans. Percept. Psychophys. 61, 1089–1101.
- Castro, L., Kennedy, P.L., Wasserman, E.A., 2010. Conditional same-different discrimination by pigeons: acquisition and generalization to novel and few-item displays. J. Exp. Psychol. Anim. Behav. Process. 36, 23–28.
- Castro, L., Wasserman, E.A., 2011. The dimensional nature of same-different discrimination behavior in pigeons. J. Exp. Psychol. Anim. Behav. Process. 37, 361–367.
- Castro, L., Wasserman, E.A., Young, M.E., in press. Variations on variability: effects of display composition on same-different discrimination in pigeons. Learn. Behav.
- Cook, R.G., Beale, K., Koban, A., 2011. Velocity-based motion categorization by pigeons. J. Exp. Psychol. Anim. Behav. Process. 37, 175–188.
- Cook, R.G., Katz, J.S., Cavoto, B.R., 1997. Pigeon same-different concept learning with multiple stimulus classes. J. Exp. Psychol. Anim. Behav. Process. 23, 417–433.
- Cook, R.G., Roberts, S., 2007. The role of video coherence on object-based motion discriminations by pigeons. J. Exp. Psychol. Anim. Behav. Process. 33, 287–298.
- Cook, R.G., Shaw, R., Blaisdell, A.P., 2001. Dynamic object perception by pigeons: discrimination of action in video presentations. Anim. Cogn. 4, 137–146.
- Cook, R.G., Wasserman, E.A., 2012. Relational discrimination learning in pigeons. In: Zentall, T.R., Wasserman, E.A. (Eds.), Oxford Handbook of Comparative Cognition. Oxford University Press, New York.
- Dittrich, W.H., Lea, S.E.G., Barrett, J., Gurr, P.R., 1998. Categorization of natural movements by pigeons: visual concept discrimination and biological motion. J. Exp. Anal. Behav. 70, 281–299.
- Driver, J., McLeod, P., Dienes, Z., 1992. Motion coherence and conjunction search: implications for guided search theory. Percept. Psychophys. 51, 79–85.
- Emmerton, J., 1986. The pigeons' discrimination of movement patterns (Lissajous figures) and contour-dependent rotational invariance. Perception 15, 573–588.
- Goto, K., Lea, S.E.G., 2003. Discrimination of direction of movements in pigeons following previous experience of motion/static discrimination. J. Exp. Anal. Behav. 80, 29–42.
- Goto, K., Lea, S.E.G., Dittrich, W.H., 2002. Discrimination of intentional and random motion paths by pigeons. Anim. Cogn. 5, 119–127.
- Hodos, W., Smith, L., Bonbright Jr., J.C., 1975. Detection of the velocity of movement of visual stimuli by pigeons. J. Exp. Anal. Behav. 25, 143–156.
- Jitsumori, M., Natori, M., Okuyama, K., 1999. Recognition of moving video images of conspecifics by pigeons: effects of individuals, static and dynamic motion cues and movement. Anim. Learn. Behav. 27, 303–315.
- Jitsumori, M., Wright, A.A., Cook, R.G., 1988. Long-term proactive interference and novelty enhancement effects in monkey list memory. J. Exp. Psychol. Anim. Behav. Process. 14, 146–154.
- Katz, J.S., Wright, A.A., Bachevalier, J., 2002. Mechanisms of same/different abstractconcept learning by rhesus monkeys (*Macaca mulatta*). J. Exp. Psychol. Anim. Behav. Process. 28, 358–368.
- Kingstone, A., Bischof, W.F., 1999. Perceptual grouping and motion coherence in visual search. Psychol. Sci. 10, 151–156.
- Koban, A., Cook, R.G., 2009. Rotational object discrimination by pigeons. J. Exp. Psychol. Anim. Behav. Process. 35, 250–265.
- Martinoya, C., Delius, J.D., 1990. Perception of rotating spiral patterns by pigeons. Biol. Cybern. 63, 127–134.

Mulvanny, P., 1978. Velocity discrimination by pigeons. Vision Res. 18, 531–536. Nakamura, N., Ushitani, T., Jitsumori, M. Do pigeons perceptually organize objects

moving in unison? An investigation using a visual search task, in preparation. Neiworth, J.J., Rilling, M.E., 1987. A method for studying imagery in animals. J. Exp. Psychol. Anim. Behav. Process, 13, 203–214.

- Nguyen, A.P., Spetch, M.L., Crowder, N.A., Winship, I.R., Hurd, P.L., Wylie, D.R.W., 2004. A dissociation of motion and spatial-pattern vision in the avian telencephalon: implications for the evolution of visual streams. J. Neurosci. 24, 4962–4970.
- Pisacreta, R., 1982. Stimulus control of the pigeons' ability to peck a moving target. J. Exp. Anal. Behav. 37, 301–309.
- Rilling, M.E., LaClaire, T.L., 1989. Visually guided catching and tracking skills in pigeons: a preliminary analysis. J. Exp. Anal. Behav. 52, 377–385.
- Royden, C.S., Wolfe, J.M., Klempen, N., 2001. Visual search asymmetries in motion and optic flow fields. Percept. Psychophys. 63, 436–444.
- Sands, S.F., Wright, A.A., 1980. Serial probe recognition performance by a rhesus monkey and a human with 10- and 20-item lists. J. Exp. Psychol. Anim. Behav. Process. 6, 386–396.
- Santiago, H.C., Wright, A.A., 1984. Pigeons memory: same/different concept learning, serial probe recognition acquisition, and probe delay effects on the serialposition function. J. Exp. Psychol. Anim. Behav. Process. 10, 498–512.
- Siegel, R.K., 1970. Apparent movement detection in the pigeon. J. Exp. Anal. Behav. 14, 93–97.
- Wasserman, E.A., Frank, A.J., Young, M.E., 2002. Stimulus control by same-versusdifferent relations among multiple visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 28, 347–357.
- Wasserman, E.A., Hugart, J.A., Kirkpatrick-Steger, K., 1995. Pigeons show same/different conceptualization after training with complex visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 21, 248–252.

- Wasserman, E.A., Young, M.E., 2010. Same-different discrimination: the keel and backbone of thought and reasoning. J. Exp. Psychol. Anim. Behav. Process. 36, 3–22.
- Wasserman, E.A., Young, M.E., Nolan, B.C., 2000. Display variability and spatial organization as contributors to the pigeon's discrimination of complex visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 26, 133–143.
- Wertheimer, M., 1923. Untersuchungen zur Lehre von der Gestalt, II. Psychol. Forsch. 4, 301–350.
- Wilkie, D.M., 1986. Some factors affecting pigeons' visual tracking behavior. Behav. Process. 12, 287–297.
- 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. Anim. Learn. Behav. 16, 436–444.
- Wright, A.A., Santiago, H.C., Sands, S.F., 1984. Monkey memory: same/different concept learning, serial probe acquisition, and probe delay effects. J. Exp. Psychol. Anim. Behav. Process. 10, 513–529.
- Young, M.E., Wasserman, E.A., Ellefson, M.R., 2007. A theory of variability discrimination: finding differences. Psychonom. Bull. Rev. 14, 805–822.
- Young, M.E., Wasserman, E.A., Garner, K.L., 1997. Effects of number of items on the pigeon's discrimination of same from different visual displays. J. Exp. Psychol. Anim. Behav. Process. 23, 491–501.
- Young, M.E., Wasserman, E.A., Hilfers, M.A., Dalrymple, R., 1999. The pigeon's variability discrimination with lists of successively presented visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 25, 475–490.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Humans deploy diverse strategies in learning same-different discrimination tasks

Leyre Castro*, Edward A. Wasserman

Department of Psychology, The University of Iowa, United States

ARTICLE INFO

Article history: Received 9 July 2012 Received in revised form 14 September 2012 Accepted 27 September 2012

Keywords: Same-different learning Individual strategies Entropy

ABSTRACT

Prior research suggests that variability discrimination is basic to same-different conceptualization (Young and Wasserman, 2001). In that research, people were trained with 16-item arrays; this training might have encouraged people to use perceptual variability to solve the task. Here, two groups of participants were trained with either 2- or 16-item Same and Different arrays (Groups 2 and 16, respectively). Participants had to learn which of two arbitrary responses was correct for the arrays without being told about the "sameness" or "differentness" of the stimuli. Surprisingly, 52% of participants in Group 2 did not learn the discrimination compared to only 21% of participants in Group 16; also, learners in Group 16 reached higher accuracy levels sooner and their choice responding was faster than learners in Group 2. A large disparity in the variability (measured by entropy) between the Same and Different arrays evidently helped participants to learn the same-different task. As well, in Group 16, we found the same two patterns of performance-Categorical and Continuous-as in prior research (Castro et al., 2006; Young and Wasserman, 2001). In Group 2, we again found the Categorical cluster, but we lost the genuine Continuous cluster and we observed a novel strategy: some participants developed a highly inclusive notion of "sameness" that applied to any array containing at least two identical icons. These findings indicate that individuals may deploy a multiplicity of possible strategies when learning a seemingly simple same-different discrimination.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

To classify sets of identical or nonidentical items as *same* or *dif-ferent* requires some level of abstract conceptualization because this classification must be based on the *relation* between or among the presented items, thereby transcending the individual items themselves. Humans learn to classify collections of items as same or different at a very early age, but very little is known about the mechanisms and intricacies involved in how this relational learning unfolds (but see, for example, Sagi et al., 2012). As well, scant attention has been paid to possible different strategies that may be deployed in human' acquisition of the same–different concept.

In the realm of animal cognition, however, distinguishing among the different strategies that animals may use to discriminate same from different items has been essential to deciding whether or not mastering a same-different discrimination involves relational learning. Thus, in the pursuit of what constitutes unequivocal concept formation, researchers including Wright et al. have found that

E-mail address: leyre-castroruiz@uiowa.edu (L. Castro).

animals, specifically pigeons, exhibit a variety of learning strategies (e.g., item-specific learning, restricted-domain relational learning, or true relational learning) when presented with a same-different discrimination task (Elmore et al., 2009). Studying these different strategies may help to shed light on the processes and mechanisms underlying learning of this fundamental capacity in both humans and animals.

Young and Wasserman (2001) first observed striking and surprising strategic disparities when they began to explore how university students solve a same-different discrimination task involving complex visual displays. In their study, participants were told to observe a series of arrays and to learn which of two arbitrary responses was correct for each of the arrays. Some arrays comprised 16 copies of the same icon, whereas other arrays comprised 16 distinctly different icons (Fig. 1, bottom). Importantly, participants were not instructed to report "same" or "different" to the visual displays; instead, they were merely told that they had to learn to press the correct response button. Participants were provided with feedback in the form of a tone (correct) or a flash of the screen (incorrect), but the words same and different were never used in the instructions nor were participants provided with any information that could have directed them toward any particular feature of the displays.

^{*} Corresponding author at: Department of Psychology, E11 Seashore Hall, The University of Iowa, Iowa City, IA 52242, United States.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.09.015



Fig. 1. Examples of the 2-icon and 16-icon Same and Different arrays used in this and prior (Castro et al., 2006; Young and Wasserman, 2001) experiments. In the current experiment, Group 2 was trained with 2-icon arrays and Group 16 was trained with 16-icon arrays.

After learning to make one response to 16-icon Same arrays and a second response to 16-icon Different arrays, participants were presented with arrays containing different mixtures of identical and nonidentical icons. Mixture arrays were shown to participants on infrequent nondifferentially reinforced probe trials. Fig. 2 illustrates seven different points along the continuum of mixture types. Participants were also tested with arrays comprising fewer than 16 icons; these testing arrays could contain 2, 4, 8, 12, or 14 items which were either the same as or different from one another. It was expected that when participants were given mixtures of same and different icons, they would respond "different" so long as any of the items differed from one another-after all, that is what different seems to mean to most of us. As well, varying the number of icons should have no effect on humans' discrimination behavior; 2 different items are just as different from one another as are 16 different items.

The results were utterly unexpected. The participants' choice responding to the testing arrays that contained fewer than 16 icons or that were mixtures of same and different icons revealed dramatically disparate profiles: 80% of the participants fell into one cluster and 20% fell into the second. The discrimination behavior of participants in the Continuous cluster (20%) was adversely affected by reducing the number of depicted icons, but only on Different trials. These Continuous participants were more likely to choose "same" to the Different arrays as the number of icons was reduced: "different" responses to 2-, 4-, 8-, 12-, and 14-icon Different arrays averaged 12%, 19%, 50%, 85%, and 90% correct, respectively. In other words, these participants strongly reported "same" to the 2- and 4-icon Different arrays; but, they consistently chose "same" to the Same arrays regardless of the number of icons. The Continuous participants also exhibited strong sensitivity to the full range of display variability in the Mixture arrays; as the mixture was changed from mostly same to mostly different icons, responding progressively changed from mostly "same" to mostly "different". So, participants in the Continuous cluster responded as if display variability fell along a continuous scale.



Fig. 2. Examples of Mixture arrays portraying different levels of entropy.

Participants in the Categorical cluster (80%) behaved dramatically differently, treating the discrimination more categorically. These Categorical participants were largely unaffected by the number of icons in the Same and Different arrays. Plus, when any of the icons were different on Mixture trials, Categorical participants tended to choose "different;" only when all of the icons were identical did they strongly choose "same."

If we set aside the fact that all of the items in the Same arrays are identical and that all of the items in the Different arrays are nonidentical, then we can appreciate that the Same arrays and the Different arrays actually represent the two ends of a *continuum of variability*. Same arrays represent minimal variability, with all 16 icons the same as one another; Different arrays represent maximal variability, with all 16 icons different from one another; and Mixture arrays represent intermediate degrees of variability. To quantify variability—or entropy—one can use the following equation originally proposed by information theorists Shannon and Weaver (1949):

$$H(D) = -\sum_{a \in D} p_a \log_2 p_a \tag{1}$$

where H(D) is the entropy of display D, a is a type of item in D, and p_a is the proportion of items of that type within the display. A Same array has 16 identical icons, so there is only one category with a probability of occurrence of 1.0, yielding an entropy of 0.0. Different arrays contain one occurrence of each of 16 icons or categories, yielding an entropy of 4.0. The entropy of Mixture arrays can be closer to 0.0 or to 4.0, the two endpoints of the entropy dimension.

Given this framework, it seemed that, rather than learning a qualitative same-different discrimination, the Continuous participants had learned a quantitative discrimination based on the entropy (Shannon and Weaver, 1949) instantiated by the displayed items. The response pattern of the Continuous participants was actually familiar to us. Prior to this human study, we had conducted extensive research with pigeons given same-different discrimination tasks (e.g., Wasserman et al., 1995; Young and Wasserman, 1997; Young et al., 1997). The discrimination task for pigeons was basically the same as that described for people: pigeons were taught to peck one report button when they viewed a 16-item Same array and to peck a second report button when they viewed a 16-item Different array. Once the pigeons reached the learning criterion, we explored the effects of intermediate degrees of variability (Young and Wasserman, 1997). In three experiments, the pigeons showed a smooth transition in "same" and "different" responding as the Mixture arrays changed from all same to all different. The entropy metric nicely fit the data. Entropy could thus account for pigeons' and our Continuous human participants' responding to Mixture arrays.

In addition, pigeons showed progressive decrements in discrimination accuracy when the number of items in Same and Different arrays was reduced from the training value of 16; accuracy was especially poor with 2- and 4-item arrays (Young et al., 1997). As with humans, this decrement in accuracy was only observed on Different trials; accuracy on Same trials involving a small number of icons remained high (see also Fagot et al., 2001, for a similar decrease in accuracy when the number of items is reduced on a relational matching-to-sample task with baboons).

Notably, the notion of entropy could make sense of the peculiar pattern of behavior that we observed when we decreased the number of depicted items. Our entropy analysis suggests that, when pigeons-and even some humans-are trained to discriminate 16icon Same arrays from 16-icon Different arrays, they will have learned to make one response ("same") to arrays with minimal entropy (0.0) and a second response ("different") to arrays with maximal entropy (4.0). During testing, organisms should distribute their responses to the testing arrays as a function of entropy; arrays with entropies closer to 0.0 should be more likely to be classified as "same," whereas arrays with entropies closer to 4.0 should be more likely to be classified as "different." The entropy of a 2-item Different array, 1.0, is thus more similar to that of 16-item Same arrays, 0.0, than it is to that of 16-item Different arrays, 4.0. Entropy discrimination should therefore prompt classification of 2-item Different arrays as "same" rather than "different," in accord with our empirical findings. These results thus represent an important and counterintuitive confirmation of different species' use of entropy in this task.

In Young and Wasserman (2001), participants were always trained with 16-icon arrays. Their testing performance might have worsened either because of the *decrease* in the number of icons or because of the *change* in the number of icons from the training value—a possible case of stimulus generalization decrement.

If the generalization decrement account were true, then training with 2, 4, 8, 12, and 16 icons from the outset should eliminate any effect of the number of items, because all of these quantities would be equally familiar. But, if training with displays of diverse numbers of icons still produced poor performance with displays of smaller numbers of icons, then this result would attest to real discrimination difficulty with small numbers of items.

Castro et al. (2006) thus pursued humans' discrimination of multiple-item arrays by again teaching university students to make discriminative report responses to identical and nonidentical arrays. But, here, the training arrays did not always involve 16 items; instead, the training arrays randomly involved 2, 4, 8, 12, or 16 items. After participants had learned the task, we focused on their choice responding to the Mixture testing arrays. We again observed two strikingly disparate patterns of discrimination behavior. Most participants (70%) treated the Mixture testing displays categorically with the remaining participants (30%) responding continuously. Participants in the Categorical cluster responded "different" when any of the icons in the Mixture arrays were different; only when all of the icons were identical did they strongly report "same." By contrast, Continuous participants responded as if display variability fell along a continuous dimension: as the Mixture arrays were changed from including mostly same icons to including mostly different icons, these participants' responding changed from mostly "same" reports to mostly "different" reports.

So, in two studies, our same-different discrimination task divulged two disparate patterns of performance: Categorical and Continuous. In the present study, we explored factors that might lead to these different performance patterns. We trained one group of participants with 2-icon Same and Different arrays (Group 2) and a second group of participants with 16-icon Same and Different arrays (Group 16); then, we tested all of the participants with 16icon Mixture arrays and with arrays containing different numbers of items, as in the studies described above.

It might be the case that 16-icon discrimination training encourages participants to rely on variability to solve the task. If this were so, then we should again find a substantial cluster of Continuous participants in Group 16; however, we might find very few if any participants in Group 2 falling into the Continuous cluster because of the small disparity in entropy between Same and Different arrays involving only 2 items.

As well, it might be the case that Continuous participants represent a small subset of the entire population that is generally disinclined to respond relationally. In order to see if there might be differences in general higher-order cognitive abilities between individuals in the Categorical and Continuous clusters, we administered a version of the Raven Advanced Progressive Matrices Test (Raven, 1989) to our participants; this test is commonly deemed to measure general cognitive processing (Court and Raven, 1982) rather than cognitive processing that is peculiar to any specific task.

2. Method

2.1. Participants

A total of 255 introductory psychology students at the University of Iowa received course credit for their voluntary participation. Participants were pseudo-randomly assigned to Group 2 (n = 158) and to Group 16 (n = 97).¹

¹ Because incoming data indicated that the attrition rate was going to be much higher in Group 2 than in Group 16, we increased the number of participants in Group 2 in order to have an equal number of valid subjects in each of the two groups.

2.2. Visual Stimuli

The particular displays that participants were shown on each trial comprised several distinctive black-and-white items chosen from a set of 24 training icons; a novel set of 24 icons was used to create transfer testing trials with the same number of items as the training trials, but with unfamiliar icons. For any given Same array, a single icon from the set was randomly chosen and was used to make up displays of 2 or 16 icons, for Groups 2 and 16, respectively. For any given Different array, 2 or 16 icons from the set were selected with no repetitions (see Fig. 1 for examples of these arrays). These methods were suitably modified when it was necessary to display 4, 8, or 12 icons.

The icons were distributed over 25 locations arrayed in a 5×5 grid, thereby leaving 9 or more locations blank. The central matrix position was used in all of the arrays; the rest of the icons were placed immediately adjacent to one another (vertically, horizon-tally, or diagonally). In addition, each of the 1 cm \times 1 cm icons could be positioned in any part of each of the 2 cm \times 2 cm cells. This distribution procedure resulted in disordered stimulus arrays in which vertical or horizontal alignment of the icons was precluded (Wasserman et al., 2002).

2.3. Procedure

2.3.1. Training

The participants were seated at one of four identically configured iMac computers. The program to run this experiment was developed in MatLab[®] with Psychtoolbox extensions (Brainard, 1997; Pelli, 1997; http://psychtoolbox.org/). The instructions to participants read as follows:

You will be observing a series of displays and attempting to learn which response is correct for each display. Your goal is to accurately predict the correct response for each display. At the beginning, you will see a white square with a black cross in the center. You should click once anywhere on the white square when you are ready to begin. Once you do so, a display will appear on the screen. You will have to choose one of two possible responses: blue or orange square. If you choose the correct button, then you will hear a pleasant tone, and you will be moved on to the next display. If you choose the incorrect button, then the display will appear again until you choose the correct button. When you are making your choice response, please respond as quickly as possible while still being accurate.

No information was provided that could have directed the participant toward any particular aspect of the displays. As well, at no point in the instructions were the words *same* or *different* used. Once each participant indicated an understanding of the procedure, the experimenter started all of the programs for that contingent of participants.

Each trial began with a white square containing a black cross in the center. The participant initiated a trial by mouse-clicking on this white square, which presented either a Same or a Different array. After 1 s, two report keys—one blue square and one orange square—appeared to the right and left of the display. The blue square was the correct response for Same arrays and the orange square was the correct response for Different arrays. The blue square was on the left and the orange square was on the right for half of the participants; the reverse was true for the other half of the participants. A correct choice response was followed by a pleasant high-pitch chime, whereas an incorrect choice response was followed by an unpleasant low-pitch buzz and a message on the screen saying: "Try again." The participant was presented with the same trial until the correct response was made. These correction trials were not scored in data analyses. The training period

Table 1

Statistical attributes of the Same, Different, and Mixture arrays.

Type of array	Icon arrangement	Entropy	Number of icon types
Same	2 icons	0.0	1
	4 icons	0.0	1
	8 icons	0.0	1
	12 icons	0.0	1
	16 icons	0.0	1
Different	2 icons	1.0	2
	4 icons	2.0	4
	8 icons	3.0	8
	12 icons	3.6	12
	16 icons	4.0	16
Mixture	14a-2b	0.5	2
	*8a-8b	1.0	2
	*12a-3b-1c	1.0	3
	*13a-1b-1c-1d	1.0	4
	11a-1b-1c-1d-1e-1f	1.5	6
	4a-4b-4c-4d	2.0	4
	8a-1b-1c-1d-1e-1f-1g-1h-1i	2.5	9
	*2a-2b-2c-2d-2e-2f-2g-2h	3.0	8
	*5a-3b-1c-1d-1e-1f-1g-1h-1i-1j	3.0	10
	*6a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k	3.0	11
	4a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k-1l-1m	3.5	13

Note: In the top half of the table are the Same and Different training arrays. In the bottom half of the table are the Mixture testing arrays, which always comprised 16 icons. Each letter designates a randomly chosen icon type that was different from those designated by the other letters in a string. The numeral preceding a letter indicates that the icon type occurred that specific number of times. The * designates Mixture arrays with identical entropy levels (1.0 or 3.0), but created with a different number of icon types.

comprised 48 trials: 4 randomized blocks of 12 Same and Different trials each.

2.3.2. Testing

After the 48 training trials, the testing period began. The session continued without a noticeable change, but testing arrays were randomly interspersed among the training arrays at a relatively low rate. During testing, 3 randomized blocks of 53 trials were given consisting of 32 Same and Different training arrays, two 16-icon (for Group 16) or 2-icon (for Group 2) Same and Different arrays created from a new icon set, 8 Same and Different arrays with different numbers of icons (4, 8, 12, and 16, for Group 2; 2, 4, 8, and 12 for Group 16), and 11 different Mixture arrays, one each with entropy levels of 0.5, 1.5, 2.0, 2.5, and 3.5, and three each of 1.0 and 3.0. The Mixture arrays always comprised 16 icons. The particular mixtures that were used for producing these entropy scores are shown in Table 1. Fig. 2 depicts some examples of these Mixture arrays.

We used three different methods for generating the 16-icon Entropy 1.0 and Entropy 3.0 arrays (see Fig. 3) as a manipulation check to ensure control by entropy rather than by other properties of the displays (e.g., the number of icon types; see Young and Wasserman, 2001). In Method 1, each type of icon appeared an equal number of times, so that when entropy was 1.0, there were 2 icon types presented 8 times each and when entropy was 3.0, there were 8 icon types presented 2 times each. In Method 2, one of the types of icons appeared three times, one other was allowed to appear more than once (12 times when entropy was 1.0 and 5 times when entropy was 3.0), whereas the others could each appear only once. And, in Method 3, one of the types of icons was allowed to appear more than once, whereas the others could each appear only once, so that when entropy was 1.0, one icon type was presented 13 times and 3 icon types were presented just once each, and when entropy was 3.0, one icon type was presented 6 times and 10 icon types were presented just once each.



Fig. 3. Examples of Entropy 1.0 and Entropy 3.0 arrays created with three different methods (see text for details).

Differential feedback continued for trials involving Same and Different training arrays; a pleasant high-pitch chime was presented if the response was correct, whereas an unpleasant low-pitch buzz and a "Try again" message were presented followed by repetition of the trial if the response was incorrect. For testing arrays, there were no correct or incorrect responses; "correct" feedback was always given, so that participants would always hear the high-pitch chime and advance to the next trial regardless of their responses. In this way, repeated testing could be conducted without teaching the participants any particular response to these arrays.

2.3.3. Raven Advanced Progressive Matrices Test

Once the same-different computer task ended, participants completed Set I and the short form of Set II (Arthur and Day, 1994) of the Raven Advanced Progressive Matrices (APM) Test. Set I contains 12 items; the short form of Set II contains one third (12 items) of the original number of items (36), selected so that they maintain the overall progressive difficulty of the long form.

The original Raven APM Test, as well as its shortened version, consists of a series of homogenous, but progressively more difficult problem items that require the participant to choose, from eight

options, the one that best completes a displayed pattern presented across a matrix of three rows and three columns. An exemplary problem is illustrated in Fig. 4.

No normative data are currently available for the short form of the Raven APM Test. Nonetheless, it was not our purpose to compare our sample to a reference population, but rather to see if this test can predict the different clusters of participants that we find in this task. The average completion time for the short form is approximately 15 min.

In all of the reported tests of statistical significance, an alpha level of .05 was adopted. We used the Tukey HSD test for comparisons between independent measures and paired-samples *t*-tests for comparisons between dependent measures; when needed, the Bonferroni correction was used.

3. Results

We chose an inclusion criterion of 70% correct choices on Same and Different trials during the last training block; any participant failing to meet this criterion was dropped from subsequent analyses. We adopted this inclusion criterion in order to obtain the clearest view of discriminative performance during testing. Of the 255 participants, 152 met this criterion; we eliminated 82 out of



Fig. 4. Examples of one of the problems in the Raven Advanced Progressive Matrices (APM) Test. A 3 × 3 incomplete matrix is presented on top. Participants have to choose, from the eight options provided below, the one that best completes the matrix.

158 (52%) participants in Group 2 and 21 out of 97 (21%) participants in Group 16. We discuss this rate of attrition below.

3.1. Training

3.1.1. Choice accuracy

During training, accuracy rapidly rose in both Groups 2 and 16, as Fig. 5 (top) shows; also, the increase in accuracy seemed to be faster in Group 16 than in Group 2. A 2 (Group 2 vs. Group 16) \times 4 $(training block) \times 2$ (type of trial: Same vs. Different) mixed analysis of variance (ANOVA) on the percentage of correct responses revealed a main effect of group, F(1,150) = 37.95, p < .001; overall accuracy was lower in Group 2 (80%) than in Group 16 (87%). There was also a main effect of training block, F(3,450) = 147.17, p < .01, confirming that overall accuracy rose from 66% in Block 1 to 98% in Block 4. The Group × Block interaction was significant as well, F(3,450) = 2.66, p < .05. Tukey HSD post hoc comparisons ($\alpha = .05$) revealed that accuracy in Group 16 was higher than in Group 2 in Block 1 (71% and 61%, respectively) and Block 2 (88% and 78%, respectively), but not in Block 3 or Block 4. By the end of training (Block 4), accuracy was very similar in Group 16 (99%) and Group 2 (97%).

There were no important disparities between Same and Different trials. Overall accuracy was 84% on Same trials and 84% on Different trials, so the main effect of type of trial type was not significant. There was a Group × Type of Trial interaction, F(1,150) = 2.66, p < .05, due to overall accuracy on Same trials being slightly lower than on Different trials in Group 16 (86% vs. 89%, respectively), but accuracy on Same trials being slightly higher than on Different trials in Group 2 (81% vs. 79%, respectively). Nonetheless, there were no significant disparities between Same and Different trials in the last block of training neither in Group 2, t(75) = 0.44, p > .10, nor in Group 16, t(75) = 1.53, p > .10.

3.1.2. Reaction time

Reaction time (RT) progressively fell in both Groups 2 and 16. as shown in Fig. 5 (bottom): also, Group 16 exhibited shorter RTs than Group 2, especially at the end of training. A 2 (Group 2 vs. Group $16) \times 4$ (training block) $\times 2$ (type of trial: Same vs. Different) mixed ANOVA on mean log(RT) scores (we transformed RTs in ms into natural logarithms to normalize the data) revealed a main effect of group, F(1,150) = 17.96, p < .001; overall, participants in Group 16 (M = 604 ms) responded with shorter RTs than participants in Group 2 (M = 679 ms). There was also a main effect of training block, F(3,450) = 172.99, p < .001, confirming that overall RT fell from Block 1 (M = 1063 ms) to Block 4 (M = 438 ms). The Group × Block interaction was significant as well, F(3,450) = 4.59, p < .01. Tukey HSD post hoc comparisons revealed that, at the beginning of training, participants in Group 2 and Group 16 exhibited similar RTs (M = 1041 ms and M = 1085 ms, respectively); however, by the end of training (Block 4), participants in Group 16 (M = 408 ms) responded with significantly shorter RTs than participants in Group 2 (M = 515 ms).

There was also a main effect of type of trial, F(1,150)=8.19, p < .01, because overall RT on Same trials (M = 616 ms) was shorter than RT on Different trials (M = 667 ms). Nonetheless, this disparity was not significant in the last block of training (M = 440 ms and M = 477 ms, for Same and Different trials, respectively), t(150) = 1.55, p > .10.

3.2. Testing

3.2.1. Choice accuracy

3.2.1.1. Mixture arrays. Our prior studies revealed substantial individual differences in same–different discriminative performance (Castro et al., 2006; Young and Wasserman, 2001). In order to explore the possibility of disparate response profiles in our current experiment, we performed Ward's (1963) hierarchical cluster



Fig. 5. Top, mean percentage of correct responses in Group 2 and Group 16 throughout the 4 blocks of training. Bottom, mean logarithmic transformation of reaction time (RT) in Group 2 and Group 16 throughout the 4 blocks of training. Error bars indicate the standard error of the means.

analysis on participants' percentage of "different" responses to the Mixture arrays in each of the two groups. We first focused on the highest-level two-cluster division in the dendrogram. In Group 16, we found two distinct patterns of performance that closely reproduced the two response profiles—Categorical (74% of the participants) and Continuous (26% of the participants)—that we had discovered in our prior studies (Castro et al., 2006; Young and Wasserman, 2001).

Participants in the Continuous cluster of Group 16 exhibited strong sensitivity to the full range of display variability in the Mixture arrays (Fig. 6, top left); as the mixture of icons was changed from mostly same icons to mostly different icons, these participants' responding changed from mostly "same" to mostly "different." Thus, participants in the Continuous cluster responded as if sameness/differentness was measured along a continuous dimension—variability.

Participants in the Categorical cluster behaved dramatically differently. When presented with Mixture arrays, they completely ignored the various levels of variability and treated the discrimination categorically: only when all of the icons were identical did they consistently report "same"; when any of the icons was different, these participants reported "different" (Fig. 6, top left). Thus, we called these participants Categorical-D (D for different). So, Group 16 disclosed the same two patterns of responding as observed in our prior work.

However, this was not the case for Group 2, in which the distinctive response profiles were better represented by a three-cluster division (see Fig. 6, top right). We did find Categorical participants, most of whom (69%) behaved exactly as the Categorical-D cluster in Group 16: when any of the icons in the Mixture arrays were different, these participants reported "different," and only when all of the icons were identical did they report "same." Curiously, a small percentage of participants in Group 2 (10%) also responded in a categorical fashion, but in precisely the opposite way: as long as at least 2 of the items in the Mixture arrays were identical, these participants reported "same"; only when all of the items were different did they tend to respond "different." Thus, we called these individuals Categorical-S (S for same) and we reserved the Categorical-D designation to refer to the more numerous categorical participants.

Finally, a third cluster of Group 2 participants (21%) superficially seemed to qualify as Continuous, but their pattern of responding to the Mixture arrays (see Fig. 6, top right) was a bit peculiar compared to Continuous participants in Group 16 (see Fig. 6, top left). These Group 2 Continuous participants' "different" reports rose as entropy increased from 0.0 to 1.5. But, when entropy was 2.0, "different" reports plummeted; these participants now deemed Entropy 2.0 arrays to be "same." For the next entropy level, 2.5, these participants' responses were mostly "different" and they continued to be so for all higher levels of entropy.

To confirm and further examine these observations, we conducted a 2 (Group 16 vs. Group 2) \times 2 (cluster: Categorical-D vs. Continuous)² \times 9 (entropy level: 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0) mixed ANOVA on the percentage of "different" responses on Mixture trials. There was a main effect of cluster, F(1,142) = 1115.5, p < .001, and a main effect of entropy, F(8,1136) = 461.28, p < .0001. There was also a Group × Cluster interaction, F(1,142) = 4.35, p < .05, suggesting disparities in the performance of the two clusters depending on the group, a Group \times Entropy interaction, *F*(8,1136) = 20.57, *p* < .001, suggesting disparities in how entropy affected each of the groups, and a Cluster × Entropy interaction, F(8,1136) = 110.30, p < .001, suggesting disparities in how entropy affected each of the clusters. Finally, the three-way Group × Cluster × Entropy interaction was significant as well, F(8,1136) = 21.29, p < .001, due to disparities in the patterns of responding of Continuous participants in Groups 2 and 16 (see Fig. 6, top).

The Categorical-D clusters in Groups 2 and 16 were almost identical, regardless of the participants having been trained with arrays containing 2 or 16 items. However, the Continuous clusters in Groups 2 and 16 proved to be quite disparate. Although Continuous participants in Group 16 showed the typical effect of entropy—with the number of "different" responses gradually increasing as the level of entropy increased (confirmed by linear trend analysis; F(1,142) = 809.62, p < .001)—the pattern of responding of Continuous participants in Group 2 was decidedly more

² Because we had two clusters of participants in Group 16, but three clusters of participants in Group 2, a between-group ANOVA with cluster as a factor would be incomplete. Thus, in order to be able to perform an ANOVA between groups, we excluded from this first analysis (and from subsequent analyses involving between-group comparisons) the Categorical-S cluster in Group 2; performance of the Categorical-S cluster was later analyzed in comparison with the other two clusters in Group 2.



Fig. 6. Top, mean percentage of "different" responses to the Mixture arrays in Group 2 and Group 16. Bottom, mean logarithmic transformation of reaction time (RT) to the Mixture arrays in Group 2 and Group 16. The response patterns are separately portrayed, in Group 16, for the Categorical-D and Continuous participant clusters and, in Group 2, for the Categorical-D, Categorical-S, and Continuous participant clusters. Entropy 0.0 (Same arrays) and Entropy 4.0 (Different arrays) contain all 16 same and all 16 different icons, respectively. In Group 16, the 16-icon Same and Different arrays were training arrays, whereas in Group 2 they were testing arrays; they are included in the graphs for comparison purposes. Error bars indicate the standard error of the means.

irregular. The percentage of "different" responses rapidly increased from Entropy 0.0 (2%) and 0.5 (8%) to Entropy 1.0 (51%) to Entropy 1.5 (79%), then it dramatically dropped for Entropy 2.0 (16%), but rose again for Entropy 2.5 (91%), decreased for Entropy 3.0 (79%), and rose once more for Entropy 3.5 (100%) and Entropy 4.0 (96%). So, although the different levels of entropy seemed to be having an effect on these participants' responding (there was indeed an overall tendency for the percentage of "different" responses to rise as entropy increased), it did not seem to be the same effect that we had seen in earlier research and that we saw here in Group 16; some other feature of the Mixture arrays appeared to be playing a role in the performance of the Continuous participants in Group 2.

In order to better understand the pattern of responding of Continuous participants in Group 2, we examined the Mixture arrays in greater detail (see Fig. 3 and Table 1). The Entropy 2.0 array—the one that most clearly failed to support the progressively incremental pattern of responding—contained 4 icons presented 4 times each: that is, 4 groups of all-same stimuli. The Continuous cluster in Group 2 judged this array to be "same" (84%). In addition, looking at arrays with entropy levels of 1.0 and 3.0 was especially interesting because these displays had been created using divergent methods, so that we had three different types of Entropy 1.0 and Entropy 3.0 arrays, but composed from different numbers and frequencies of icon types. If Continuous participants' responses were controlled by entropy and no other properties of the displays, then there should have been no disparities among the three types of arrays sharing the same entropy level.

Overall responding of Continuous participants in Group 2 to the Entropy 1.0 and Entropy 3.0 arrays did not depart so obviously from the incremental Continuous pattern; however, when we looked at responding to each of the three types of Entropy 1.0 and Entropy 3.0 arrays (see Fig. 3), we found unexpected disparities. When the Entropy 1.0 array was created using Method 1 (each type of icon involved an equal number of items)—so that it contained 2 icons presented 8 times each or 2 groups of all-same stimuli (8a-8b array)-participants reported the array to be "same" (94%); when the Entropy 1.0 array contained other combinations that lacked even numbers of same icons (created with Methods 2 and 3), participants tended to report "different" (71% for the 12a-3b-1c array and 77% for the 13a-1b-1c-1d array). Similar disparities were observed for Entropy 3.0 arrays. When the Entropy 3.0 array was created using Method 1-so that it contained 8 icons presented 2 times each or 8 groups of all-same stimuli (2a-2b-2c-2d-2e-2f-2g-2h array)-participants had a slight tendency to say "same" (54%); when the Entropy 3.0 array contained other combinations that lacked even numbers of same icons (created with Methods 2 and 3), participants gave clear "different" reports (93% for the 5a-3b-1c-1d-1e-1f-1g-1h-1i-1j array and 98% for the 6a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k array).

These striking disparities within the same entropy level were not observed in the Continuous cluster in Group 16. For the Entropy 1.0 arrays, the percentage of "different" responses was 18%, 33%, and 28% for the 8s8s, 12s3s1s, and 13s3d arrays, respectively. For the Entropy 3.0 arrays, the percentage of "different" responses was 91%, 78%, and 83% for the 2a-2b-2c-2d-2e-2f-2g-2h, 5a-3b-1c-1d-1e-1f-1g-1h-1i-1j, and 6a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k arrays, respectively.

Disparities were significant between the 8a-8b array and the other two Entropy 1.0 arrays in the Continuous cluster of Group 2 [t(15) = 5.39, p < .001, and t(15) = 6.24, p < .001, for the 12a-3b-1c array and the 13a-1b-1c-1d array, respectively], but not in the Continuous cluster of Group 16 [t(19)=2.26, p>.05, and t(19)=1.30,p > .05, for the 12a-3b-1c array and the 13a-1b-1c-1d array, respectively]; as well, disparities between the 2a-2b-2c-2d-2e-2f-2g-2h array and the other two Entropy 3.0 arrays were significant in the Continuous cluster of Group 2 [t(15) = 4.58, p < .001, and t(15) = 3.62, p < .001, ap < .01, for the 6a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k array and the 5a-3b-1c-1d-1e-1f-1g-1h-1i-1j array, respectively], but not in the Continuous cluster of Group 16[t(19) = 1.75, p > .05, and t(19) = 1.63, pp > .05, for the 6a-1b-1c-1d-1e-1f-1g-1h-1i-1j-1k array and the 5a-3b-1c-1d-1e-1f-1g-1h-1i-1j array, respectively]. The absence of significant disparities in Group 16 supports entropy as the dimension controlling the same-different reports of these Continuous participants. Thus, the responses of Continuous participants in Group 16 seemed to be controlled by entropy, whereas the responses of the Continuous participants in Group 2 seemed to be controlled by other properties of the displays, such as the number of times that each type of icon appeared; so, their pattern of responding might not properly be termed Continuous.

As noted above, we identified a third cluster of participants in Group 2 whose performance did not correspond with the performance of any other cluster found in our prior studies. These Categorical-S participants responded "same" as long as 2 items in the arrays were the same as one another. In a separate analysis, we compared the three clusters of participants in Group 2. A 3 (cluster: Categorical-S vs. Categorical-D vs. Continuous) \times 9 (entropy: 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0) ANOVA on the percentage of "different" responses to Mixture arrays yielded a main effect of cluster, F(2,73) = 765.47, p < .001, a main effect of entropy, F(8,584) = 95.03, p < .001, and a Cluster × Entropy interaction, *F*(16, 1168)=45.63, p < .001, confirming that variations in entropy affected the three clusters in a divergent ways. Categorical-S participants were truly categorical; when presented with Mixture arrays, they completely ignored the various levels of variability and treated all of the Mixture arrays in the same way. However, unlike the more numerous contingent of Categorical-D participants, Categorical-S gave special status to "differentness;" only when variability was maximal did they tend to report that an array was "different."

3.2.1.2. Number arrays. Next, we examined the effect of changing the number of items in Same and Different arrays. Both groups were presented with arrays containing 2, 4, 8, 12, and 16 items—either all same or all different. So, in Group 16, we can see the effect of decreasing the number of items from the training value of 16; and, in Group 2, we can see the effect of increasing the number of items from the training the number of items from the training value of 2.

As Fig. 7 (top) illustrates, changing the number of items affected each of the groups and clusters differently. In Group 16, Categorical-D participants were unaffected by the number of icons in the Same and Different arrays. Continuous participants' responding to the Same arrays was unaffected as well, but their accuracy to Different arrays progressively fell as the number of icons in the arrays decreased. Therefore, Categorical-D and Continuous participants in Group 16 showed the same pattern of performance as we had reported before (Castro et al., 2006; Young and Wasserman, 2001). In Group 2, only Categorical-S participants were affected by a change in the number of icons. Curiously, their drop in accuracy was in the opposite direction as the Continuous cluster in Group 16: as the number of items increased, their accuracy progressively fell to Different arrays while their responding to Same arrays was unaffected.

A 2 (Group 2 vs. Group 16) × 2 (cluster: Categorical-D vs. Continuous) × 2 (type of array: Same vs. Different) × 5 (number: 2, 4, 8, 12, 16) mixed ANOVA on the percentage of correct responses revealed that all of the main effects and interactions were significant. Most important, the four-way Group × Cluster × Type of Array × Number interaction was significant as well, F(4,568) = 5.58, p < .001, testifying to the differential effects of the number of items on Same and Different trials for the Categorical-D and Continuous clusters in Group 16, but not in Group 2 (Fig. 7, top).

In a separate analysis, we compared the three clusters of participants in Group 2. A 3 (cluster: Categorical-D vs. Categorical-S vs. Continuous) × 2 (type of array: Same vs. Different) × 5 (number: 2, 4, 8, 12, 16) repeated measures ANOVA on the percentage of correct responses revealed a main effect of cluster, F(2,73) = 32.64, p < .001, a main effect of type of array, F(1,73) = 16.43, p < .001, and a Cluster × Type of Array interaction, F(2,146) = 19.62, p < .001. The three-way Cluster × Type of Array × Number interaction was not significant (F < 1), despite the clearly disparate pattern of performance on Different trials for the Categorical-S participants' accuracy fell as the number of icons was increased in Different arrays, F(1,73) = 7.04, p < .01. None of the other linear trend analyses was significant (all ps > .10).

As was true in our prior studies, the pattern of responding of the Continuous cluster in Group 16 to arrays displaying different numbers of items (Fig. 7, top left) can be straightforwardly explained by the notion of entropy: when people are trained to discriminate 16-icon Same arrays from 16-icon Different arrays, they learn to discriminate minimal entropy (0.0) from maximal entropy (4.0). Entropy for Same arrays is always 0.0, regardless of the number of items in the display, whereas entropy for Different arrays decreases as the number of items is reduced. When Continuous participants are shown Same arrays containing fewer items, their accuracy remains high because the entropy value for those arrays is identical to the entropy training value. But, when Continuous participants are shown Different arrays containing fewer than 16 items, their accuracy drops because, as the number of items is reduced, the entropy values for those arrays progressively deviate from the 16-item Different entropy training value and progressively approximate the 16-item Same entropy training value.

The Continuous cluster of participants in Group 2 showed no change in accuracy when the number of items was increased. This insensitivity to number might be considered to be consistent with the entropy account as well. These participants were trained to discriminate 2-icon Same arrays from 2-icon Different arrays; so, they learned to discriminate minimal entropy (0.0) from a higher, but similar level of entropy (1.0). When these participants were presented with Different arrays containing a larger number of items, their accuracy would not be expected to decline because higher levels of entropy ought to be even easier to discriminate from 0.0 than the entropy training value of 1.0.

3.2.1.3. Novel-item arrays. Next, we examined choice accuracy in each of the groups and clusters to testing arrays with the same number of items as in the training phase, but created from a new and different pool of items. These trials are critical for determining the generality of the participants' same-different discrimination behavior and for claims that this behavior is conceptual in nature.

In Group 2, a 3 (cluster: Categorical-D vs. Categorical-S vs. Continuous) \times 2 (type of items: familiar vs. novel) ANOVA on the



Fig. 7. Top, mean percentage of "different" responses to the Same and Different arrays with different numbers of icons in Group 2 and Group 16. Bottom, mean logarithmic transformation of reaction time (RT) to the Same and Different arrays with different numbers of icons in Group 2 and Group 16. The response patterns are separately portrayed, in Group 16, for the Categorical-D and Continuous participant clusters and, in Group 2, for the Categorical-D, Categorical-S, and Continuous participant clusters. Entropy 0.0 (Same arrays) and Entropy 4.0 (Different arrays) contain all 16 same and all 16 different icons, respectively. In Group 16, the 16-icon Same and Different arrays were training arrays, whereas in Group 2 they were testing arrays. Error bars indicate the standard error of the means.

percentage of correct responses to 2-item arrays revealed a main effect of cluster, F(2,73)=4.85, p<.01, due to overall accuracy in the Categorical-S cluster (89%) being slightly lower than in the Categorical-D (97%) and Continuous (99%) clusters, as disclosed by Tukey HSD post hoc comparisons. There was no main effect of type of items or an interaction, so none of the clusters of participants in Group 2 were detrimentally affected by the presentation of novel items (96% and 96% for familiar and novel items, respectively).

In Group 16, a 2 (cluster: Categorical-D vs. Continuous) × 2 (type of items: familiar vs. novel) ANOVA on the percentage of correct responses to 16-item arrays revealed a main effect of cluster, F(1,74) = 13.16, p < .001, due to overall accuracy in the Categorical-D cluster (99%) being slightly higher than in the Continuous (94%) cluster. There was also a main effect of type of items, F(1,74) = 13.16, p < .01, and a Cluster × Type of Items interaction, F(1,74) = 5.88, p < .05. Presentation of new items had no effect on the Categorical-D cluster's accuracy (99% and 98% for familiar and novel items, respectively), t(55) = 1.62, p > .10, but it did have a small detrimental effect

on the Continuous cluster's accuracy (98% and 90% for familiar and novel items, respectively), t(19) = 2.20, p < .05. Thus, although they responded with quite high accuracy to the novel testing items, the Continuous participants in Group 16—who were sensitive to the different degrees of perceptual variability in the arrays—were somewhat sensitive to the particular icons that had been presented in the training arrays.

3.2.2. Reaction time

3.2.2.1. Mixture arrays. We next examined participants' RTs to the Mixture testing arrays. In Group 16, we observed a pattern that was very similar to that seen in our prior studies (Castro et al., 2006; Young and Wasserman, 2001). For Categorical-D participants, RTs initially rose from Entropy 0.0 to Entropy 0.5 arrays, followed by a steady fall as entropy declined from 0.5 to 4.0. For Continuous participants, RTs rose from Entropy 0.0 to Entropy 1.5 arrays, followed by a steady fall as entropy declined from Entropy 1.5 to Entropy 4.0 (Fig. 6, bottom left).

In Group 2, Categorical-D participants showed a similar pattern to Categorical-D participants in Group 16. Categorical-S participants responded differently, with longer RTs and RTs which steadily rose from Entropy 0.0 to Entropy 4.0 arrays. Continuous participants generally exhibited the longest RTs and, although it could be argued that their RT pattern was similar to the Continuous-D participants in Group 16—with an increase from Entropy 0.0 to Entropy 1.5 arrays and a decrease from Entropy 1.5 to Entropy 4.0 arrays—their profile also contained some sharp irregularities (Fig. 6, bottom right).

A 2 (Group 16 vs. Group 2) × 2 (cluster: Categorical-D vs. Continuous) × 9 (entropy level: 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0) mixed ANOVA on mean log(RT) scores revealed that all main effects and interactions (except for the three-way Group × Cluster × Entropy interaction) were significant. Overall, Group 2 responded with longer RTs (M=514 ms) than Group 16 (M=440 ms), F(1,142)=197.48, p <.001. The Group × Cluster interaction revealed that, although Categorical-D participants responded with shorter RTs than Continuous participants in both Groups 2 and 16 (as disclosed by Tukey HSD post hoc comparisons), the RT disparity between Categorical-D and Continuous participants was smaller in Group 16 (M=428 ms and M=477 ms, respectively) than in Group 2 (M=399 ms and M=1162 ms).

The increase in RT from Entropy 0.0 to 0.5 was significant for Categorical-D participants (M = 411 ms and M = 592 ms, for Entropy 0.0 and Entropy 0.5, respectively) in Group 16, t(55) = 7.30, p < .001; but, although numerically in the same direction, it was not significant in Group 2 (M = 455 ms and M = 503 ms, for Entropy 0.0 and Entropy 0.5, respectively), t(51) = 1.74, p > .05. Linear trend analysis disclosed that, from Entropy 0.5 to Entropy 4.0, the decline in RT was significant in both Categorical-D clusters, F(1, 74) = 83.94, p < .001, and F(1, 66) = 15.81, p < .001, for Group 16 and Group 2, respectively.

Trend analysis revealed a significant quadratic component in the RT pattern of Continuous participants in Group 16, F(1, 74) = 17.73, p < .01 (as evidenced by the steady increase in RT from Entropy 0.0 to Entropy 1.5, and the steady decrease from Entropy 1.5 to Entropy 4.0, see Fig. 6, bottom left), and in Group 2 as well, F(1, 66) = 11.69, p < .01 (for which RT increased from Entropy 0.0 to Entropy 2.0, and then decreased from Entropy 2.0 to Entropy 4.0, see Fig. 6, bottom right).

We next analyzed the performance of the Categorical-S cluster in relation to the Categorical-D and Continuous clusters in Group 2. A 3 (cluster: Categorical-D vs. Categorical-S vs. Continuous) × 9 (entropy level: 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0) mixed ANOVA on mean log(RT) scores revealed a main effect of cluster, F(2,73) = 237.66, p < .001. Tukey HSD post hoc comparisons disclosed that Categorical-D participants (M=399ms) responded with shorter RTs than participants in the other two clusters; Categorical-S participants (M=800 ms) responded with longer RTs than Categorical-D participants, but they responded with shorter RTs than Continuous participants, who exhibited the longest RTs of all (M=1162 ms). There was also a main effect of entropy, F(8,584) = 4.89, p < .05, and a Cluster × Entropy interaction, F(16,1168) = 4.17, p < .001, revealing that changes in the entropy level of the arrays affected the three clusters of participants differently: as entropy increased from 0.5 to 4.0, Categorical-D participants exhibited shorter RTs, whereas Categorical-S participants exhibited longer RTs, and Continuous participants exhibited an irregular RT pattern, suggesting once more that Continuous participants' responses to Mixture arrays were not based on entropy, but rather on some more inefficient strategy (most likely related to the specific number or proportion of type of icons in the array) that made them the slowest responding participants of all.

Linear trend analysis confirmed that the increase in RTs from Entropy 0.0 to Entropy 4.0 in the Categorical-S cluster was significant, F(1, 73) = 19.33, p < .01. So, Categorical-S participants appear to have been adversely affected by entropy: as entropy increased, they became slower to respond, presumably because, as variability increased, it became more difficult to detect matching items in the arrays (see Fig. 2).

3.2.2.2. Number arrays. We next examined participants' RTs to Same and Different arrays that comprised different numbers of icons. In Group 16 (Fig. 7, bottom left), all of the participants responded with longer RTs as the number of items was decreased. Even Categorical-D participants were sensitive to changing the number of items in the arrays, especially on Different trials, as we had found in prior experiments (Castro et al., 2006; Young and Wasserman, 2001).

In Group 2 (Fig. 7, bottom right), Categorical-D participants' RTs were the shortest of all and they did not appear to be strongly affected by variations in the number of items. Continuous participants' responding was affected, but in the opposite direction from Categorical-D and Continuous participants in Group 16; their RTs rose as the number of items was increased. Categorical-S participants—who generally exhibited the longest RTs—also showed a rise in RT as the number of items increased, but only on Different trials.

In order to confirm the statistical significance of these observations, we performed a 2 (Group 16 vs. Group 2) × 2 (cluster: Categorical-D vs. Continuous) × 2 (type of array: Same vs. Different) × 5 (number: 2, 4, 8, 12, 16) mixed ANOVA on mean log(RT) scores. This analysis revealed a main effect of group, F(1,145)=37.89, p < .001, a main effect of cluster, F(1,145)=76.25, p < .001, a Group × Cluster interaction, F(1,145)=74.08, p < .001, a Group × Number interaction, F(1,145)=12.49, p < .001, a Group × Number interaction, F(4,580)=26.73, p < .001, and a Cluster × Number interaction, F(4,580)=7.01, p < .001.

Most important, the three-way $\text{Group} \times \text{Cluster} \times \text{Number}$ interaction was significant, F(4,580) = 7.16, p < .001, confirming that the RT patterns of Categorical-D and Continuous clusters were affected by the number of items in discrepant ways in Groups 2 and 16. No other main effects or interactions were significant.

Linear trend analysis confirmed that, in Group 16, both Categorical-D and Continuous participants' RTs rose as the number of items was decreased; this trend was significant on both Same and Different trials [F(1, 74) = 9.67, p < .01 and F(1, 74) = 9.96, p < .01 on Same trials for Categorical-D and Continuous participants, respectively, and F(1, 74) = 65.74, p < .001 and F(1, 74) = 13.14, p < .01 on Different trials for Categorical-D and Continuous participants, respectively]. According to entropy, a reduction in the number of items should slow responding, but only on Different trials. Increasing RTs as the number of items was decreased—especially on Same trials—might be due to generalization decrement from the training value of 16.

Next, we compared the three clusters of participants in Group 2. A 3 (cluster: Categorical-D vs. Categorical-S vs. Continuous) × 2 (type of array: Same vs. Different) × 5 (number: 2, 4, 8, 12, 16) repeated measures ANOVA on mean log(RT) scores revealed a main effect of cluster, F(2,146)=93.37, p < .001, a main effect of type of array, F(1,73)=20.34, p < .001, a main effect of number, F(4,292)=9.48, p < .001, a Cluster × Type of Array interaction, F(2,146)=20.61, p < .001, a Cluster × Number interaction, F(8,584)=5.46, p < .001 and, most important, a three-way Cluster × Type of Array × Number interaction, F(8,584)=2.01, p < .05, confirming that, in Group 2, the Categorical-D, Categorical-S, and Continuous clusters' RT patterns were differently affected by the number of items.

Linear trend analysis disclosed that Categorical-D participants' RTs did not vary significantly depending on the number of items, F(1,73) = 0.13, p > .10, but Continuous and Categorical-S

participants' RTs did. Indeed, Continuous participants' RTs rose as the number of items was increased on both Same and Different trials [F(1,73) = 23.66, p < .001 and F(1,73) = 25.31, p < .001, respectively], both tendencies at variance with entropy; this directional disparity further suggests that these participants were not truly Continuous. Finally, linear trend analysis disclosed that Categorical-S participants' RTs significantly rose as the number of items increased, but this trend was significant only on Different trials, F(1,73) = 20.34, p < .001.

Categorical-S participants' RT patterns on Number arrays nicely complements their accuracy data; these participants were quick and accurate on Same trials, regardless of the number of items, but they were increasingly prone to errors and were slower to respond on Different trials as the number of items in the arrays progressively deviated from the training value of 2. It seems that these participants could easily generalize the concept of "sameness" that they had learned during training with 2-item arrays, but they found it increasingly difficult to extend the concept of "differentness" beyond arrays containing 2 different items.

3.2.2.3. Novel item-arrays. Finally, we examined RTs in each of the groups and clusters to testing arrays comprising the same number of items as in the training phase, but created from a new and different pool of items.

In Group 2, a 3 (cluster: Categorical-D vs. Categorical-S vs. Continuous) × 2 (type of items: familiar vs. novel) ANOVA on mean log(RT) scores to 2-item arrays revealed a main effect of cluster, F(2,73) = 6.86, p < .01, due to overall RT in the Categorical-S cluster (M = 573 ms) being longer than in the Categorical-D (M = 411 ms) and Continuous (M = 429 ms) clusters, as disclosed by Tukey HSD post hoc comparisons. There was no main effect of type of items (M = 422 ms and M = 440 ms, for familiar and novel items, respectively) and no interaction, so none of the clusters of participants' RTs in Group 2 were affected by the presentation of novel items. In Group 16, a 2 (cluster: Categorical-D vs. Continuous) × 2 (type of items: familiar vs. novel) ANOVA on mean log(RT) scores to 16item arrays revealed no main effects or interactions (M = 376 ms and M = 386 ms, for familiar and novel items, respectively).

Thus, none of the participants' RTs, neither those in Group 2 nor those in Group 16, were affected by the introduction of arrays containing novel items.

3.3. Clusters performance during training

Our three clusters were identified by participants' responses to the Mixture testing arrays. It is possible that those clusters might have exhibited differential *acquisition* performance as well. Because we previously documented disparities in accuracy and RT during training between *groups* given different numbers of items, we now focus on disparities in accuracy and RT among the *clusters*.

A one-way ANOVA on mean accuracy of the five different clusters (Categorical-D and Continuous in Group 16; Categorical-D, Categorical-S, and Continuous in Group 2) revealed a main effect of cluster, F(1,147) = 5.56, p < .05. Tukey HSD post hoc comparisons revealed that only accuracy of the Categorical-S cluster (68%) was significantly lower than accuracy of the other four clusters (88% and 82% for Categorical-D clusters in Group 16 and Group 2, respectively, and 86% and 82% for Continuous clusters in Group 16 and Group 16 and Group 2, respectively).

A one-way ANOVA on mean $\log(RT)$ scores of the five different clusters revealed a main effect of cluster, F(1,147)=4.35, p < .05. Tukey HSD post hoc comparisons revealed that only RTs of the Categorical-S cluster (M = 996 ms) were significantly slower than RT of the other four clusters (M = 593 ms and M = 665 ms for Categorical-D clusters in Group 16 and Group 2, respectively, and



Fig. 8. Mean scores in the short form of the Raven Advanced Progressive Matrices (APM) Test for each of the clusters in Group 2 and Group 16, and for the nonlearners in Group 2 and Group 16. Above each column is noted the number of participants in each of the conditions. The maximum score, when all problems are solved correctly, is 24. Error bars indicate the standard error of the means.

M=636 ms and *M*=601 ms for Continuous clusters in Group 16 and Group 2, respectively).

So, the Categorical-S participants in Group 2 proved to be the least accurate and slowest of all participants. Importantly, neither accuracy nor RT in training could anticipate the disparities between the Categorical-D and Continuous clusters.

3.4. Raven APM Test

The Raven APM Test is scored by simply summing the number of correctly solved problems. We added the scores for the Set I (12 problems) and Set II (12 problems) forms in order to create a combined score; the maximum score was thus 24. Mean combined scores, including learners and nonlearners, were very similar in Group 2 (M = 17.12, SE = 0.30) and Group 16 (M = 17.37, SE = 0.41), so the disparity between groups in the percentage of nonlearners (52% in Group 2 compared to 21% in Group 16) and learning rate (higher accuracy and lower RTs in Group 16 than in Group 2 during training) was not due to preexisting disparities in higher-order general cognitive ability between the groups.

Inspection of Fig. 8 reveals that the Raven scores of nonlearners, in both groups, were lower than that those of learners. Although the Raven scores for the Categorical participants were somewhat higher than for the Continuous participants, there did not appear to be large disparities in performance among those Group 2 and Group 16 participants who learned the discrimination.

Because of the lack of disparities between Categorical-D and Categorical-S participants in Group 2, we combined their Raven scores for analysis. A 2 (Group 16 vs. Group 2) × 3 (cluster: Categorical vs. Continuous vs. Nonlearners) ANOVA on the mean combined Raven scores yielded a main effect of cluster, F(2, 506)=3.93, p < .05. Tukey HSD post hoc comparisons revealed that the scores of nonlearners (M=16.31, SE=0.41) were significantly lower than the scores of Categorical (M=18.00, SE=0.32) and Continuous (M=17.25, SE=0.69) participants; there was no discrepancy between Categorical and Continuous participants. There was no effect of group or an interaction between the factors.

3.5. Nonlearners

It is noteworthy that many of our university students failed to learn the task to criterion (70% correct on Same and Different trials) by the end of the training phase. Out of a total of 255 participants, only 152 met this criterion. This large attrition rate has precedents in our laboratory. In Young and Wasserman (2001), where participants were trained with 16-item arrays, 7% of the participants did not reach the learning criterion (75% correct in the final training block); however, in Castro et al. (2006), where training arrays randomly comprised 2, 4, 8, 12, or 16 icons, the percentage of nonlearners rose to 38% (even when we used a slightly more lenient criterion of 70% over the entire experiment). In the current study, 21% of the participants (n = 21) in Group 16 did not learn by the end of the training phase; strikingly, 52% of the participants in Group 2 did not learn by the end of the training phase (n = 82). Clearly, learning to discriminate Same from Different 2-item arrays—without any instructions about the relevant properties of the arrays—proved to be a very difficult task for our students.

The 70% learning criterion in our experiment had to be met by the last block of training before the testing phase began, so that we could properly evaluate the effects of presenting Mixture arrays and arrays containing different numbers of items. However, the testing period itself contained a relatively high proportion of Same and Different training arrays, so participants might still be learning during the testing phase. Indeed, of the 21 participants in Group 16 who did not meet criterion by the end of training, 13 of them (62%) did so by the last block of testing; only 8 participants (38%) were still at chance level at the end of testing. However, the percentage of late learners was much lower in Group 2; 22 participants (27%) met criterion by the last block of testing, but 60 participants (73%) never did so. It seems that even extended training would still have yielded a very large attrition rate in Group 2. The intriguing question is why so many of these participants failed to solve our same-different discrimination task

When we realized that many students were not meeting our discrimination learning criterion, we included a questionnaire at the end of the experiment, asking participants to "specify what strategy you followed in order to choose between the blue and the orange squares." The nonlearners whom we queried (all of them from Group 2) gave answers like: "I tried to look for patterns," or "I looked for some form of arrow pointing to either the blue or the orange" (one of the icons was an arrow and another icon was a finger pointing, but none of the other 22 icons had any pointing characteristic), or "I looked at which pictures kept appearing on either the right or left side" (note that the icons were randomly located in the arrays so, by chance, they could be located on the right or left portion of the display, closer to one response button than to the other). Consistent with the deployment of these idiosyncratic and unreliable strategies is the fact that nonlearners were much slower to respond than learners; during training, the mean RT was 641 ms (SD = 22) for learners and 892 ms (SD = 21) for nonlearners.

Although we introduced the questionnaire late in the experiment and we cannot properly analyze the content of these limited responses, these comments suggest that participants were deploying item-specific strategies based on visual properties of the icons in the arrays rather than attending to the relationship between or among the icons. Accordingly, nonlearners' Raven APM Test scores were significantly lower than learners' scores, suggesting that nonlearners were people who were generally less likely to use relational information to solve cognitive problems.

4. Discussion

We undertook the present research in order to expand our understanding of humans' same-different categorization behavior. Of key interest was the performance of humans on same-different discrimination tasks involving either 2 or 16 items.

Perhaps the most remarkable of our results was that half (52%) of our university students did not learn a nonverbal same-different discrimination task when it involved arrays containing only 2 items

(Group 2). We are quite confident that all of these nonlearning participants would have quickly mastered the task if the instructions had explicitly mentioned that they had to distinguish between "same" and "different" pairs of items (nonetheless, this remains an empirical question and, possibly, a control group to include in future studies). However, it proved to be very difficult for participants to extract the "sameness" and "differentness" of the displays without those instructions. When the same–different task involved arrays containing 16 items (Group 16), the percentage of nonlearners was much lower (21%), clear evidence that increasing the disparity in variability or entropy between the Same and Different arrays helped participants to apprehend the relationship among the items and, therefore, to learn the same–different task.

Some of our nonlearners' comments in the post-experimental questionnaire suggest that their responding was controlled by the presence of specific items or the spatial arrangement of items in the display. Thus, these participants were deploying an ineffective low-level perceptual strategy rather than a high-level conceptual strategy in their attempts to solve the task. Accordingly, the nonlearners' scores in the Raven APM Test, a measure of higher-order cognitive ability, were reliably lower than the scores of participants who did learn the task.

Yet, even among the participants who learned the discrimination, there was a clear advantage when training involved 16-item arrays. Group 16 reached higher accuracy levels sooner than Group 2 (Fig. 5, top), and choice responding in Group 16 entailed shorter RTs than in Group 2 (Fig. 5, bottom). The benefit of using a large number of items to learn a same-different discrimination can readily be explained by entropy. When arrays contain 16 items, there is a large disparity in entropy between Same (Entropy 0.0) and Different (Entropy 4.0) arrays, so these arrays are easily discriminated. But, when arrays contain only 2 items, the disparity in entropy between Same (Entropy 0.0) and Different (Entropy 1.0) arrays is small, so these arrays are more difficult to discriminate. Thus, an entropy account offers a straightforward explanation for the large percentage of nonlearners in Group 2 compared to Group 16 as well as the lower training accuracy and longer RTs in Group 2 compared to Group 16.

Another matter of interest concerned how training with low (Group 2) or high (Group 16) entropy disparity levels would affect responding to other arrays containing different numbers of icons (Number arrays) or mixtures of same and different icons (Mixture arrays). In Group 16 (as in prior studies: Castro et al., 2006; Young and Wasserman, 2001), performance on Mixture trials revealed two disparate patterns of discrimination behavior. People's predominate inclination (74% of our participants) was to treat the arrays categorically, with the remainder (26% of our participants) responding continuously. Participants in the Categorical-D cluster responded "different" when any of the icons were different; only when all of the icons were identical did they report "same" (Fig. 6, top left). Accordingly, their RTs were short with Entropy 0.0 (Same) arrays; they rose sharply with Entropy 0.5 arrays (where the category boundary between same and different presumably lies); and, they became progressively shorter as they approached the opposite categorical endpoint with the Entropy 4.0 (Different) arrays (Fig. 6, bottom left). Hence, Categorical-D participants' RTs to the Mixture arrays disclosed that they had the greatest difficulty discriminating displays with entropies near the category boundary. Prior studies involving physical dimensions such as size or length have also found that RTs increase as response uncertainty increases near categorical boundaries (Ashby et al., 1994).

By contrast, Continuous participants in Group 16 responded as if display variability fell along a continuous dimension: as the Mixture displays were changed from including mostly same icons to mostly different icons, these participants' responding gradually changed from mostly "same" reports to mostly "different" reports (Fig. 6, top left). Continuous participants' RTs to these Mixture arrays rose from Entropy 0.0 to Entropy 1.5, followed by a steady fall to Entropy 4.0, suggesting that the decision boundary between low and high variability was located near Entropy 1.5 (Fig. 6, bottom left).

When arrays containing fewer items were presented to participants in Group 16, Categorical-D participants' accuracy remained very high on both Same and Different trials; on the contrary, Continuous participants' accuracy dropped, but only on Different trials (Fig. 7, top left). A decrease in the number of items translates into a decrease in the entropy of Different arrays, but not of Same arrays. Consequently, accuracy for Different arrays containing fewer items declines, because the entropy value is reduced in relation to the training entropy value. Interestingly, RTs for both clusters of participants became longer when the number of items was reduced on both Same and Different trials (Fig. 7, bottom left). Entropy would predict lengthening RTs on Different trials, but not on Same trials; so, it could be that this overall lengthening reflects a generalization decrement due to the change in the number of icons from the training value, rather than to a sensitivity to entropy per se. However, Castro et al. (2006) trained participants with arrays containing different numbers of items (2, 4, 8, 12, and 16) from the outset and also observed a progressive increase in RT as the number of items in the arrays decreased. Even if generalization decrement is involved in this slowing as the number of items decreases, it cannot explain all of our findings, which suggests a role for entropy as well.

In Group 2, performance on Mixture trials revealed that the majority of people (69%) also treated the discrimination categorically (Fig. 6, top right), responding almost exactly as the Categorical-D cluster in Group 16 (Fig. 6, top left). Interestingly, the Categorical-D cluster in Group 2 showed little sensitivity to entropy, even in their RTs to arrays containing different numbers of items, as was true for Categorical-D participants in Group 16. So, this cluster of participants was the only subgroup for which variability did not seem to matter.

However, training with 2-item arrays did not generate a typical Continuous cluster. Although we have used the term "Continuous" to designate 21% of participants in Group 2, their pattern of responding did not fully conform to the usual Continuous pattern. When presented with Mixture arrays, these participants' "different" responses did not gradually increase as the mixture of icons progressively changed from mostly same to mostly different; rather, their pattern of responding was quite irregular (Fig. 6, top right) and seemed to be controlled by other properties of the displays, such as number or proportion of icon types. These "pseudo-Continuous" participants exhibited no change in accuracy when the number of items was increased (Fig. 7, top right); this sustained high level of accuracy would have been expected even if they were using entropy to solve the task, because the disparity in entropy between Same and Different arrays grows larger when arrays contain more items, thereby making the task even easier. However, these participants exhibited increasing RTs as the number of items increased (Fig. 7, bottom right), so the same-different discrimination did not really appear to become easier for them when the number of icons in the arrays was increased. Such slower RTs also suggest that entropy alone may not have been controlling these participants' performance.

Quite peculiarly, we found that a small percentage of people (10%) in Group 2 treated the discrimination categorically, but in an unexpected way. When presented with Mixture arrays, as long as 2 icons were the same as one another, these Categorical-S participants responded "same," and only when all of the items were different did they respond "different" (Fig. 6, top right). Their RTs to Mixture arrays were affected by entropy, but in the opposite direction from what entropy would predict, so that the more variable the arrays, the longer they took to make a choice response (Fig. 6, bottom right). Also, their accuracy to arrays containing larger

numbers of items dropped, but only on Different trials (Fig. 7, top right), and their RTs gradually rose as the number of items increased, again only on Different trials (Fig. 7, bottom right).

The fact that this Categorical-S pattern was characteristic of such a small minority of our participants (8 out of a total of 152 learners) is consistent with the Young and Wasserman (2002) study, which found it easier for people (and pigeons) to discriminate arrays containing all-same items from arrays containing some degree of differentness (Entropy 0.0 vs. Entropies 0.5–4.0) than to discriminate arrays containing all–different items from arrays containing some degree of sameness (Entropy 4.0 vs. Entropies 0.0–3.5). From both Young and Wasserman (2002) and the present study, we can conclude that small disparities are very salient compared to small similarities, which in turn suggests a broad *different* category and a narrow *same* category (as shown by the majority of Categorical-D participants; also see Cook and Wixted, 1997; Smith et al., 2008); the opposite possibility is very rare, but we now know that it can nonetheless occur (Categorical-S participants).

5. Final considerations

To summarize, when we trained participants with 16-item arrays, we found the same two patterns of performance-Categorical-D and Continuous-as in our prior reports (Castro et al., 2006; Young and Wasserman, 2001). When we trained participants with 2-item arrays, we again found the Categorical-D pattern cluster; however we lost the genuine Continuous cluster. It seems that training participants to discriminate low levels of variability reduces the later influence of variability on behavior. In addition, we also discovered novel strategies that we had not envisioned at the start of the project. Some participants, even when they learned to classify pairs of items as "same" or "different," were also sensitive to specific combinations of same and different items in the Mixture arrays (the participants whom we had initially labeled Continuous). Still other participants (Categorical-S) developed a highly inclusive notion of "sameness" that applied to any array containing at least 2 same icons.

Remarkably, a procedurally minor modification in our same-different discrimination task—training with arrays containing only 2 icons instead of our customary 16 icons—yielded dramatically diverse patterns of same-different discrimination behavior by university students. Such divergent behavioral patterns suggest a panoply of possible strategies that individuals may deploy when learning a seemingly simple same-different discrimination.

It should further be noted that, regardless of this variety of strategies, all of our learners exhibited extremely high transfer (over 90%) to testing arrays created from a novel set of icons. In training, we used a 24-item set size in order to create our Same and Different arrays, a relatively small number that, nonetheless, allowed for high generalization in both Groups 2 and 16. Prior studies by Wright et al. have shown that the number of training exemplars presented to animal subjects has a strong effect on their exhibiting robust transfer to novel items (Katz and Wright, 2006; Katz et al., 2002; Wright et al., 2003). In Wright's same-different task, in which only 2 items are presented simultaneously on the screen, pigeons required a set size of 256 items in order to show full transfer, whereas monkeys required a set size of 128 items. In Castro et al. (2010), using multi-item displays like those used in the current study, pigeons showed modest transfer with a 24-item set (accuracy dropped from 93% with training arrays to 69% with novel-item arrays), but they showed virtually perfect transfer with a 72-item set (accuracy was 91% for training arrays and 87% for novel-item arrays). As Katz and Wright (2006) suggested, disparities among species in the number of exemplars required to show high transfer coupled with the fact that most

species can eventually attain a high level of transfer performance suggests a *quantitative* rather than *qualitative* differences among species.

So, according to standards commonly adopted in animal cognition research, our participants evidenced the *learning a concept*. We must then conclude that concept learning (as evidenced by extremely high transfer to novel exemplars; see, Katz et al., 2007) can nevertheless be attained by deploying a diverse array of learning strategies.

Where might all of those strategies originate? We initially suspected that quite general higher-order cognitive abilities might be responsible for participants' tendencies. But the Raven APM Test did not distinguish between our Categorical and Continuous clusters. Thus, the question of why people exhibit different patterns of behavior remains very much open.

Nonetheless, our perhaps painfully complex behavioral observations do allow us to rule decisively on at least one view of human same–different discrimination behavior. Commenting on our prior pigeon work and its relation to human cognition, Mandler (2004) opined that: "Whatever the pigeons are responding to does not appear to be anything like a human concept of same–different (p. 140)." Based on an extensive program of research spanning 10 years and including several hundred university students, we would now reply: "Which human concept?"

Acknowledgement

We thank Dhayasini Ramasamy for her help in conducting this study.

References

- Ashby, F.G., Boynton, G., Lee, W.W., 1994. Categorization response time with multidimensional stimuli. Percept. Psychophys. 55, 11–27.
- Arthur, W., Day, D., 1994. Development of a short form for the Raven Advanced Progressive Matrices test. Educ. Psychol. Meas. 54, 395–403.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433-436.

- Castro, L., Kennedy, P.L., Wasserman, E.A., 2010. Conditional same-different discrimination by pigeons: Acquisition and generalization to novel and few-item displays. Journal of Experimental Psychology: Animal Behavior Processes 36, 23–38.
- Castro, L., Young, M.E., Wasserman, E.A., 2006. Effects of number of items and visual display variability on same-different discrimination behavior. Mem. Cognit. 34, 1689–1703.

- Cook, R.G., Wixted, J.T., 1997. Same-different texture discrimination in pigeons: testing competing models of discrimination and stimulus integration. J. Exp. Psychol. Anim. Behav. Process. 23, 401–416.
- Court, J.H., Raven, J., 1982. Manual for Raven's Progressive Matrices and Vocabulary Scales (Research Supplement 2, Pt. 3, Sec. 7). H.K. Lewis, London.
- Elmore, L.C., Wright, A.A., Rivera, J.J., Katz, J.S., 2009. Individual differences: either relational learning or item-specific learning in a Same/Different task. Learn. Behav. 37, 204–213.
- Fagot, J., Wasserman, E.A., Young, M.E., 2001. Discriminating the relation between relations: the role of entropy in abstract conceptualization by baboons and humans. J. Exp. Psychol. Anim. Behav. Process. 27, 316–328.
- Katz, J.S., Wright, A.A., 2006. Same/Different abstract-concept learning by pigeons. J. Exp. Psychol. Anim. Behav. Process. 32, 80–86.
- Katz, J.S., Wright, A.A., Bachevalier, J., 2002. Mechanisms of same/different abstractconcept learning by rhesus monkeys (*Macaca mulatta*). J. Exp. Psychol. Anim. Behav. Process. 28, 358–368.
- Katz, J.S., Wright, A.A., Bodily, K.D., 2007. Issues in the comparative cognition of abstract-concept learning. Comp. Cogn. Behav. Rev. 2, 79–92.
- Mandler, J.M., 2004. The Foundations of Mind: Origins of Conceptual Thought. Oxford University Press, New York.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442.
- Raven, J., 1989. The Raven Progressive Matrices: a review of national norming studies and ethnic and socioeconomic variation within the United States. Journal of Educational Measurement 26, 1–6.
- Sagi, E., Gentner, D., Lovett, A., 2012. What difference reveals about similarity. Cogn. Sci. 36, 1019–1050.
- Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.
- Smith, J.D., Redford, J.S., Haas, S.M., Coutinho, M.V.C., Couchman, J.J., 2008. The comparative psychology of same-different judgments by humans (*Homo sapiens*) and monkeys (*Macaca mulatta*). J. Exp. Psychol. Anim. Behav. Process. 34, 361–374.
- Ward Jr., J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.
- Wasserman, E.A., Hugart, J.A., Kirkpatrick-Steger, K., 1995. Pigeons show same-different conceptualization after training with complex visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 21, 248–252.
- Wasserman, E.A., Frank, A.J., Young, M.E., 2002. Stimulus control by same versus different relations among multiple visual stimuli. J. Exp. Psychol. Anim. Behav. Process. 28, 347–357.
- Wright, A.A., Rivera, J.J., Katz, J.S., Bachevalier, J., 2003. Abstract concept learning and list-memory processing by capuchin and rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 29, 184–198.
- Young, M.E., Wasserman, E.A., 1997. Entropy detection by pigeons: response to mixed visual displays after same-different discrimination training. J. Exp. Psychol. Anim. Behav. Process. 23, 157–170.
- Young, M.E., Wasserman, E.A., 2001. Entropy and variability discrimination. J. Exp. Psychol. Learn. Mem. Cogn. 27, 278–293.
- Young, M.E., Wasserman, E.A., 2002. Detecting variety: what's so special about uniformity? J. Exp. Psychol. Gen. 131, 131–143.
- Young, M.E., Wasserman, E.A., Garner, K.L., 1997. Effects of number of items on the pigeon's discrimination of same from different visual displays. J. Exp. Psychol. Anim. Behav. Process. 23, 491–501.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Visual discrimination learning with an iPad-equipped apparatus

Kenneth J. Leising*, Joshua E. Wolf, Chad M. Ruprecht

Department of Psychology, Texas Christian University, United States

ARTICLE INFO

Article history: Received 20 August 2012 Received in revised form 9 November 2012 Accepted 26 November 2012

Keywords: Visual discrimination iPad Touchscreen Differential reinforcement of other behavior Correction method

ABSTRACT

Visual discrimination tasks are commonly used to assess visual learning and memory in non-human animals. The current experiments explored the suitability of an iPad (Apple, Cupertino, California), as a low-cost alternative touchscreen for visual discrimination tasks. In Experiment 1, rats were trained with patterned black-and-white stimuli in a successive non-match to sample procedure. Rats successfully interacted with the iPad but failed to learn to withhold responding on trials in which the sample matched the comparison. Experiment 2 used the same patterned stimuli, but the procedure was simplified to a successive discrimination procedure and we explored the use of procedures known to facilitate discrimination learning. Rats that received training with differential outcomes and a differential reinforcement of other behavior schedule successfully acquired the task. In Experiment 3, the same rats were tested in a simultaneous discrimination task and we explored the use of a correction and non-correction method during acquisition. Rats that failed to learn the discrimination in the previous experiment, improved while trained with the correction method. These experiments support the use of the iPad in visual discrimination tasks and inform future studies investigating learning and memory within a touchscreen-equipped (iPad or other) apparatus.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

In the past decade, touchscreen technology has dramatically changed the way we interact with our environment. From bank ATMs to cell phones, touchscreens improve the flexibility and ease with which we can display information and record behavior. These benefits have not been overlooked by behavioral scientists. Touchscreen displays have been used with pigeons (e.g., Allan, 1992; Blough, 1986; Pisacreta and Rilling, 1987; Wright et al., 1988), rats (e.g., Bussey et al., 1994; Cook et al., 2004; Markham et al., 1996; Sahgal and Steckler, 1994), non-human primates (e.g., Bhatt and Wright, 1992; Elsmore et al., 1989), and humans (e.g., Huguenin, 2000). Furthermore, touchscreens have been used to study the effects of neural lesions and pharmacological agents on behavior (e.g., Bussey et al., 1997a, 1997b, 1998, 2001; Parkinson et al., 1999, 2002). Though many factors have contributed to the popularity of touchscreens in research, most researchers are attracted by the ease with which visual stimuli can be created and displayed, and the flexibility to detect responses across the entire display.

Recognizing the constraints imposed by the display equipment used in research in the 1980s, Dr. Anthony Wright became one of the pioneers in developing methods for the use of touchscreen technology in research. Wright et al. (1988) suggested that the use of small numbers of exemplars in most previous reports on matchto-sample performance (a consequence of the 12-slide projectors used by most researchers at the time) encouraged item-specific, rather than relational (i.e., concept) learning. Wright et al. used a novel, horizontally mounted touchscreen-equipped display to present 152 trial-unique stimuli to a group of pigeons during match-to-sample training. When compared to a group trained with only 2 stimuli, the results revealed concept learning only in the group trained with a large number of exemplars. This experiment highlighted the benefits associated with incorporating new technology into research and demonstrated how extending the levels of an independent variable can result in gualitative differences in learning (see Wright, 2010 for a review). Wright and colleagues have also been interested in procedural changes that accelerate the development of concept learning. Wright and Delius (1994) required pigeons to dig through different colored and textured sand in match and oddity tasks and found a 100-fold acceleration in learning relative to more traditional methods. More recently, Schmidtke et al. (2010) describe how differential outcomes can accelerate the expansion of concept learning to produce better transfer to novel items in a same/different task. The present experiments are inspired by Wright and colleagues use of new technology

^{*} Corresponding author at: Department of Psychology, Texas Christian University, 2800 S. University Dr., Box 298920, Fort Worth, TX 76129, United States. Tel.: +1 817 257 6139.

E-mail address: k.j.leising@tcu.edu (K.J. Leising).

^{0376-6357/\$ -} see front matter © 2013 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.013

and manipulation of experimental parameters to better understand the nature of learning and memory. We explored the suitability of training rats in a novel iPad-equipped apparatus and explored conditions which may facilitate discrimination learning.

The use of touchscreens in research was extended to rats by Bussey et al. (1994) and Markham et al. (1996). Much of the subsequent research has investigated discrimination learning by varying object luminance (Minini and Jeffery, 2006), shape (e.g., Bussey et al., 1994, 2001, 2008; Markham et al., 1996; Minini and Jeffery, 2006; Simpson and Gaffan, 1999), or pattern (e.g., Cook et al., 2004; Prusky et al., 2002). However, a subset of the research has focused on determining methods suitable for a touchscreen environment. For example, Markham et al. (1996) reported better acquisition in rats interacting with a touchscreen when the site of reward delivery was at the opposite wall of the chamber from the touchscreen, rather than adjacent to it. Recently, Bussey et al. (2008) found that learning rate could be improved through the use of larger stimuli, longer inter-trial intervals (ITIs), and more trials per session. Lastly, Cook et al. (2004) directly compared the benefits of touchscreen technology with traditional lever press equipment. Rats were trained in a simultaneous visual discrimination task with either traditional stimulus and response equipment (e.g., lights and levers) or a touchscreen-equipped display. Rats responding to the touchscreen learned the discrimination faster than rats responding on traditional levers.

Researchers have demonstrated visual discrimination learning in rats interacting with infrared (Bussey et al., 1994, 2001, 2008; Cook et al., 2004) and pressure sensitive (Markham et al., 1996; Minini and Jeffery, 2006) touchscreens. Although both touchscreens are suitable for use with rats, the motivation to develop an alternative to the currently available infrared touchscreen systems is their cost. Assuming one already has a operant chamber for rats (e.g., test chamber, reward delivery mechanism, etc.), an additional touchscreen package will cost \$5000 (Med Associates, Georgia, VT) to \$10,000 (Lafayette Instruments, Lafayette, IN) and does not include the software (e.g., \$3000, Med Associates) or any additional software packages (e.g., autoshaping, \$1100, Lafayette Instruments). One alternative to these systems is to design a custom touchscreen chamber (see Cook et al., 2004). However, variations in the type of touchscreen purchased and the placement of the touchscreen in an operant chamber can affect learning outcomes (e.g., Markham et al., 1996). With the goal of providing a flexible, low cost alternative touchscreen system, we have recently validated the use of an iPad 2 (Apple, Cupertino, California) with rats (Wolf et al., submitted for publication).¹

The iPad provides a high-resolution display, and a capacitive touchscreen that offers drift-free stable performance. Although infrared touchscreen technology has improved and the issue of drift has been largely eliminated, the lingering issue with infrared touchscreens is that the panel frame, which houses the LEDs and phototransistors, always protrudes slightly above the screen. This becomes a problem when a rat's whiskers or other object triggers a response in lieu of a true touch. Additionally, obstructions (e.g., feces) on the thick panel frame can block the infrared sensors rendering them non-functional until the obstruction is removed. Capacitive touchscreens detect responses through changes in current as a result of direct contact with the screen and automatically recalibrate when an obstruction is present. Lastly, the iPad is wireless. This enables it to be mounted in any position and easily repositioned within an operant box or open field (e.g., on the wall or embedded in the floor). Visual Basic 6 software (Microsoft, Redmond, WA) also allows you to display and record from multiple iPads concurrently, using the same PC. Multiple iPads could be used to create a 4-walled iPad environment, or control iPads in separate locations in an open field. In sum, the use of an iPad is relatively inexpensive and the technology is reliable, durable, and enables wireless flexibility.

The current research explores the suitability of three visual discrimination procedures for use in an iPad-equipped apparatus (see Fig. 1a). Discrimination learning has long played an important role in investigations of perception and memory and remains popular among those interested in cognition and physiology (Dudchenko, 2004). Among the tasks used by researchers, successive and simultaneous discrimination procedures are most commonly used to evaluate visual learning and memory in rodents. Despite varying cognitive demands, each task requires that a subject discriminate between objects (presented successively or simultaneously) and associate the identified object with reward or non-reward. The experimental setup utilizes an iPad, operant chamber, a remote desktop program, and Visual Basic 6 software. Experiments 1-3 evaluated the iPad as a suitable device for use with rats and explored conditions which may facilitate learning. Experiment 1 tested rats in a successive match-to-sample task with patterned black-and-white images as stimuli. The same stimuli were then used in successive and simultaneous discrimination tasks in Experiments 2 and 3, respectively. Together, these experiments suggest the iPad is an attractive alternative to costly prepackaged touchscreen systems.

2. Experiment 1

In one of the first studies of memory in non-humans, Hunter (1913) evaluated delayed choice in rats, raccoons, and dogs. The animals were confined to a chamber in view of three separate choice boxes. Over one of the choice boxes a light was shown for a brief period of time. Once extinguished, the animal was freed from the chamber to select a choice box. The delay between the presentation of the light and release was then manipulated to determine the duration of working memory. In subsequent years, the use of successive presentations of stimuli in a match-to-sample format allowed a more thorough investigation into the nature of working memory. Subjects in a match-to-sample task are presented with one stimulus (sample), which is followed after some delay by a second stimulus (comparison). The subject is then required to detect whether the comparison matches the sample stimulus. Selection of the match or non-match comparison may be correct depending on the procedure selected by the experimenter. This type of procedure can be used to study the acquisition, content, and retrieval of the memory for the sample.

The purpose of the present experiment was to evaluate the use of a successive non-match to sample task with rats interacting with an iPad. Rats were presented with a sample followed by a brief delay and a second, comparison stimulus. If the comparison stimulus matched the sample, then the correct response was to withhold touching the comparison until the stimulus timed out. If the comparison did not match the sample, then a touch to the comparison stimulus was rewarded. We used four circular black-and-white patterned images as stimuli (see Fig. 1b). Previous research has shown rats can solve visual discriminations based on pattern alone (e.g., Cook et al., 2004; Lashley, 1938; Prusky et al., 2002); however, rats have been shown to rely primarily on unidimensional (e.g., size or brightness) differences in stimuli (Minini and Jeffery, 2006). Thus, we expected that the relative similarity between our stimuli would result in slow learning in the task.

2.1. Method

2.1.1. Subjects

Three female experimentally naïve Long–Evans rats (*Rattus norvegicus*) obtained from the TCU Breeding Colony served as



Fig. 1. (a) A picture of the iPad-equipped operant box. Stimuli were displayed on the wall opposite of a sucrose delivery system. (b) All four stimuli were used in Experiment 1. The two stimuli on the top row were used in Experiments 2 and 3.

subjects. Females were pair-housed in translucent plastic tubs with a substrate of wood shavings in a vivarium maintained on a 12-h dark/12-h light cycle. All experimental manipulations were conducted during the light portion of the cycle. A progressive food restriction schedule was imposed over the week prior to the beginning of the experiment, until each rat received 15 g of food each day. All animals were handled daily for 30 s during the week prior to the initiation of the study.

2.1.2. Apparatus

A test chamber measuring $30.5 \text{ cm} \times 24.1 \text{ cm} \times 29.2 \text{ cm}$ $(length \times width \times height)$ was housed in a sound- and lightattenuating environmental isolation chest (Med Associates, Georgia, VT). The side walls and ceiling of the chamber were constructed of clear Plexiglas. The front and rear walls were constructed of aluminum panels. The floor was constructed of stainless-steel rods measuring 0.5 cm in diameter, spaced 1.5 cm center-to-center. The enclosure was dimly illuminated by a 28-V, 100-mA shielded incandescent house light mounted on the top of the rear wall of the chamber, 2 cm below the ceiling. One wall of the chamber was equipped with a dipper that could deliver a sucrose solution (16%). When in the raised position, a small well (0.05 cc) at the end of the dipper arm protruded up into the feeding niche. An infrared photo-detector was positioned across the entrance to the feeding niche. When a rat placed its nose into the feeding niche to lick the sucrose solution (i.e., a nose poke), the photo beam was disrupted. The duration of sucrose access did not begin until the computer detected an interruption of the photo beam. A ventilation fan in the enclosure and a white-noise generator on a shelf outside of the enclosure provided a constant 74-dB (A-Scale) background noise.

On the wall opposite of the hopper (30.5 cm), an iPad was mounted flush against the rear of the test chamber (see Fig. 1a). The iPad features a 24.63 cm (diagonal) LED-back lit glossy widescreen display with multi-touch sensitivity. Although the iPad recognizes three different types of touches – taps, moves and gestures, only the tap function was utilized in the current experiments. The program code treated every touch as a tap by detecting and recording only touch-down (i.e., screen contact) events. For one day in pretraining, the iPad was turned off and positioned in a landscape orientation and placed at a 54° angle inside the test chamber. This procedure encouraged the rats to approach and contact (e.g., rear and lean on) the screen. On all subsequent days, the iPad was positioned in the same orientation but mounted at a 90° angle at the rear of the test chamber. The display of stimuli on the iPad, data collection, and hardware activation (houselight, dipper, and fan) were controlled by an adjacent PC. The PC to iPad connection was accomplished via a remote desktop program RDP (Mochasoft Aps, Blokhus, Denmark) downloaded from the App Store (Apple, Cupertino, California; but see also Wolf et al. for a recently developed Mac application). The programs for all of the experiments were written in Microsoft Visual Basic 6.0, which used a dynamic link-library (dll), purchased as part of the Control of Med Input/Output from Other Languages Med Associates Product (SOF-732-3), to control Med-Associates hardware.

2.2. Procedure

2.2.1. Stimuli

A 6.2 cm light gray circle served as the pretraining stimulus. During training, the stimuli were 4.7 cm circles filled with a blackand-white checker, a white on black background grid, a scattered white dot on black background, and a diagonal-striped pattern (see Fig. 1b). The training stimulus was positioned at the midpoint of the iPad, which coincided with the midpoint of the rear wall of the test chamber. The stimuli were displayed at 12.2 cm from the chamber floor.

2.2.2. Pretraining

The rats were initially trained to drink from the dipper with the iPad located in the test chamber. Sucrose was delivered for 3-s on a variable-time (VT) 60-s schedule. After rats were drinking reliably, the iPad was moved to the rear of the chamber and mounted at a 90° angle. Over the next 1–5 Days, rats were exposed to a combination of autoshaping and manual shaping. Autoshaping consisted of a 32-trial session with presentations of an 8-s pretaining stimulus followed by 3-s access to sucrose and then an 80-s inter-trial interval (ITI).

2.2.3. Successive nonmatch-to-sample training

On sessions 1–35, each discrimination training session consisted of 40 trials (20 match and 20 non-match trials). The onset of the sample stimulus was always marked by darkening of the houselight. The sample stimulus was always presented for 5 s, but the duration of the comparison stimulus was gradually reduced from 60 to 10 s across sessions. Trial termination following a touch to a comparison stimulus was constrained by a fixed-interval (FI) 1 s schedule of reinforcement. This was used initially to ensure the comparison stimulus would be viewed for a fixed amount before a response terminated the trial. The FI schedule was increased


Fig. 2. Data are from the last 35 sessions of training during Experiment 1. In Block 1, the delays were 0, 50, and 100 ms. This was increased to 0, 50, 100, 200, and 250 ms during Blocks 2–3, and then increased again to 250, 500, and 750 ms in Blocks 4–7. Error bars represent standard error of the mean.

gradually from 1 to 4s across sessions to allow a direct comparison of responding during a fixed amount of time on match and non-match trials. The delay between stimuli was also manipulated across sessions. The delay was initially chosen pseudo randomly from among 0, 50, and 100 ms. This was subsequently increased to 0, 50, 100, 200, and 250 ms; and increased again to 250, 500, and 750 ms. For all rats, a touch to a non-match comparison after the FI resulted in 3-s access to sucrose. A touch to a matching comparison after the FI terminated the trial and initiated a 16-s timeout period. A 4-s fixed-time ITI separated all trials. The houselight was off throughout the duration of a trial, but remained on during the ITI and timeout periods.

On Days 36–47, sessions consisted of 50 trials with 25 match and 25 non-match trials. A trial terminated after 10-s of the comparison stimulus or the first response following the FI-4, whichever came first. A variable-interval 500 ms (250, 500, and 750 ms) delay separated the sample and comparison stimulus. All other details were the same as described above.

2.3. Results

Fig. 2 displays data from 35 sessions of training. A percent correct for each session was calculated by dividing the number of correct responses on non-match (i.e., a touch) and match (i.e., no touch) trials by the total number of trials. As indicated by Fig. 2, there was little change across sessions and no evidence for a preference to respond on non-match trials. A statistical analysis was conducted on Days 36-47, which represented a period of consistent session parameters. A t-test against chance (50%) confirmed no subject demonstrated a preference to respond on non-match trials across the last 12 sessions of training, ts(11) < 1, ps > .05. Other measures, including a discrimination ratio (DR) calculated using responses collected within the FI-4 s period after the onset of the comparison stimulus also indicated no difference, ts(11) < 1, ps > .05. Lastly, we compared the DR during the different delay intervals (250, 500, and 750 ms) but found no preference to respond on non-match trials at any delay, *ts*(11) < 1, *ps* > .05.

All rats failed to show successive nonmatch-to-sample performance after many sessions of training. Rats may have failed to learn the correct response as a result of poor discriminability between stimuli, memory interference, or a combination. Previous research has shown that rats rely mostly on unidimensional (e.g., size or brightness) differences in the lower hemifield during simultaneous discrimination procedures (Lashley, 1938; Minini and Jeffery, 2006; Sutherland, 1961). Discrimination of patterned images can be learned but typically proceeds slowly. For example, Minini and Jeffery (2006) reported slow learning and low asymptotic performance while training rats using a touchscreen to discriminate shape. In their procedure (Experiment 3), the luminance and position of two visual stimuli were controlled for but aspect ratio, a basic property of shape, was the cue for the S+. After 40 sessions of 128 trials per session (5120 trials), mean asymptotic performance was only 64%. In their results, however, rats showed some evidence of learning in the first 10 sessions of training (1280 trials). We terminated the current experiment after the last 35 sessions (1500 trials) indicated no evidence of learning.

In addition to poor discriminability, retroactive interference as a result of short ITIs and repeated stimuli likely contributed to poor performance in this task (c.f., Frank and Wasserman, 2005; Kelly et al., 1999). For example, the rewarded comparison stimulus on one trial (i.e., a non-match trial) could become a nonrewarded comparison stimulus (i.e., a match trial) on the very next trial. Long-duration ITIs and session-unique, highly discriminable stimuli would likely facilitate acquisition of the non-match to sample task. Subsequent studies may also include a trial initiation response, which has been shown to facilitate acquisition of visual discriminations (Bussey et al., 2008), as well as further manipulate the delay between stimuli (e.g., Bussey et al., 2008; Pontecorvo, 1983; Pontecorvo et al., 1996).

3. Experiment 2

Experiment 1 showed rats can be quickly trained to interact with an iPad-equipped apparatus but fail to learn the correct response in a successive non-match to sample task. In Experiment 2, we trained four naïve rats in a successive discrimination task using two of the four cues from Experiment 1. Stimuli were designated as either S+ or S- and presented on separate trials. A target response on trials with the S+ was reinforced, whereas, a response on trials with a S- either terminated a trial with no reinforcement and a time-out. The aims of Experiment 2 were to demonstrate visual discrimination performance in an iPad-equipped apparatus with patterned stimuli, and to evaluate manipulations known to facilitate learning in tasks where generalization or response confusion may occur (e.g., Delamater et al., 2010; Lieberman et al., 1979, 1985; Urcuioli and Kasprow, 1988; Williams, 1999).

In a typical successive discrimination procedure, all trials end with the termination of the target stimulus and an empty interval (i.e., an ITI). Consequently, on S+ trials with no response and Strials with a response or not, the subject receives the same endof-trial events. The similarity in trial outcomes likely attenuates discrimination performance. In order to facilitate discrimination performance in our task, a response to the S+ was rewarded with sucrose and, for some rats a response to the S- was not rewarded and followed immediately by a flashing light of the same duration as the sucrose (c.f., Cook et al., 2004). Any facilitated learning effect would resemble the differential outcomes effect (DOE) found in instrumental (see Overmier and Linwick, 2001) and Pavlovian procedures (e.g., Delamater et al., 2010). In discriminations involving two S+ stimuli (e.g., a low and high pitched tone) and two responses (e.g., left and right lever), responding appropriately to each stimulus is enhanced if the correct responses are associated with different outcomes (i.e., O1 and O2). One theory explaining the DOE involves an "acquired distinctiveness" which develops between cues followed by different events (see Hall, 1991, for a review). In addition to acting as a unique outcome, the flashing light served to mark an incorrect response, thereby enhancing the salience of that response and discrimination learning (e.g., Lieberman et al., 1979, 1985; Urcuioli and Kasprow, 1988; Williams, 1999).

Another procedure that has been used to enhance discrimination performance is a differential reinforcement of other behavior (DRO) schedule. A DRO schedule was used successfully by Kelly et al. (1999) to train pigeons to discriminate visual stimuli in a successive discrimination task. In their procedure, if a no-go test stimulus occurred, pigeons were reinforced for withholding pecking for 5 s. If a peck did occur, the DRO timer was reset to 5 s. In the last phase of our experiment, we implemented a DRO schedule. Rats were required to withhold touching the S– for 4-s before the trial would terminate. Differential outcomes, marking procedures, and DRO schedules have all been shown to facilitate learning in discrimination tasks. We explored whether these procedural manipulations could be effective in an iPad-equipped apparatus.

3.1. Method

3.1.1. Subjects

Four experimentally naïve Long–Evans rats (*R. norvegicus*) obtained from the TCU Breeding Colony served as subjects. Two male and female rats were randomly assigned to two groups, with the constraint that one male and female was in each group. All other details are the same as in Experiment 1.

3.1.2. Apparatus

The apparatus was the same as that used in Experiment 1. However, during a session the house light could be flashed by turning the light on and off every .25 s.

3.2. Procedure

3.2.1. Stimulus displays

The stimuli were the diagonal-striped pattern and black-andwhite checkered images used in Experiment 1 (Fig. 1b, top row). The two stimuli were assigned as S+ and S-, counterbalanced across subjects. All other details are identical to that of Experiment 1.

3.2.2. Pretraining

Pretraining occurred in the same manner as described in Section 2.

3.2.3. Phase 1: discrimination training (Days 1–6)

Each discrimination training session consisted of 56 trials (28 S+ and 28 S–). A trial terminated after a touch to the stimulus or 20 s, whichever came first. The duration of the stimulus presentation was reduced from 20 to 10 s across sessions. A fixed-interval 1-s schedule (FI-1) determined whether a touch terminated a trial. The FI ensured the stimulus would be displayed for a minimum duration before a response would terminate the trial. The FI schedule was increased gradually from 1 to 4 s across sessions. For all rats, a touch to the S+ after the FI resulted in 3-s access to sucrose. For rats in the control group, a touch to the S– after the FI terminated the trial and initiated a 16-s timeout period, whereas, rats in another group were given a 3-s flashing light followed by a 16-s timeout period. All trials, correct or incorrect, were followed by a 4-s fixed-time ITI. The houselight was off during all trials, but remained on during the ITI and timeout periods.

3.2.4. Phase 2: fixed-interval discrimination training (Days 7–12)

The number of trials was increased to 100 (50 S+, 50 S-) and a fixed-interval 4-s schedule determined whether a response terminated a trial.



Fig. 3. Data from the first 4-s of S+ and S- trials in Experiment 2 were used to calculate discrimination ratios for each rat. The left-hand side of the figure displays performance during Phase 2 (sessions 7–12) and the right-hand side (light gray shading) indicates performance with the DRO schedule during Phase 3. The dotted-line patterns represent animals that received a flashing light after an incorrect response. The dotted line at .50 represents no difference between responding on S+ versus S- trials.

3.2.5. Phase 3: differential reinforcement of other behavior training (Days 13–18)

A differential reinforcement of other behavior (DRO) schedule was introduced. On this schedule, an S- trial would not terminate until 4-s elapsed with no response to the stimulus. Reinforced trials remained unchanged from Phase 2.

3.3. Results and discussion

Percent correct was calculated for all sessions in Phase 1 (M=50.08, SD=.72). Clearly, mean performance during Phase 1 indicated no preference for responding on S+ trials. In Phase 2, the number of trials increased and the FI-4 was introduced. During the FI-4, responding was uninterrupted by reward or trial termination allowing a direct comparison between S+ and S- trials. A discrimination ratio was calculated as in Experiment 1, such that a preference for the S+ was indicated by values greater than .5. Fig. 3 displays the discrimination ratio for all rats during Phases 2 and 3. A t-test against chance (.50) compared responding across all six sessions of Phase 2 and revealed a significant preference for the S+ in one rat in the group receiving the flashing light, t(5) = 3.90, p < .05. In Phase 3, the DRO procedure was implemented. The same analysis conducted on the six sessions from Phase 3 revealed that both rats in the group receiving the flashing light responded more on S+ trials than S–, ts(5)>5.33, ps <.01, whereas no rats in the control group showed this preference, ts(5) < 1, ps > .05.

These results suggest that rats can be trained to discriminate between highly similar patterned stimuli within an iPad-equipped apparatus. Evidence for successful visual discrimination was present only for the rats that received the flashing light and DRO procedures. It is beyond the scope of these data to determine whether the flashing light enhanced the salience of an incorrect response through marking or enhanced the discriminability of the S+ and S- via a differential outcomes effect. Certainly, there is support in the literature for facilitated learning as a result of both manipulations. The two rats without the differential outcome failed to learn the task. Though, the small number of subjects per group suggests caution when interpreting these results. In Experiment 3, another common task for evaluating visual learning, a simultaneous visual discrimination, was used to determine whether (1) rats that learned the successive discrimination would perform well on the first trial of a simultaneous discrimination procedure, and (2) rats that had not learned the task in Experiment 2 would acquire a simultaneous discrimination with the same stimuli.

4. Experiment 3

In Experiment 3, the two stimuli from the last experiment were presented simultaneously in positions on the left and right sides of the display. The assignment of stimulus to S+ and S- from Experiment 2 was maintained. As before, responses to the S+ were reinforced, whereas responses to the S- were non-reinforced. Acquisition of the task was evaluated by comparing responses to the S+ and S- during the first 4-s (FI) of each trial. The simultaneous discrimination task allowed subjects more of an opportunity to compare the features of the S+ and S-, which should benefit those rats that failed to learn the discrimination in Experiment 2.

We expected to observe fairly rapid learning in all rats during training with the simultaneous discrimination procedure. Bussey et al. (2008) trained naïve rats in a simultaneous discrimination with similar parameters (e.g., trials per session, ITI duration, and image size) and found evidence of learning within five sessions. In Experiment 3, we trained rats during the first three sessions with a correction method followed by another three days of training with a non-correction method. In discrimination learning paradigms, the use of a correction method permits the animal to continue to respond until a correct response is made. However, a noncorrection method indicates that a trial will terminate (or the animal will be removed from the apparatus) after an incorrect response. In Experiment 3, we explored the use of a simultaneous discrimination procedure and correction method with rats interacting with an iPad display. We expected discrimination performance to develop quickly for all rats and no decrement in responding when the non-correction method was introduced.

4.1. Method

4.1.1. Subjects

The same four Long–Evans rats (*R. norvegicus*) from Experiment 2 served as subjects.

4.1.2. Apparatus

The apparatus was the same as that used in Experiments 1 and 2.

4.2. Procedure

4.2.1. Stimuli

The stimuli were the same as those used in Experiment 2, except the diameter of the pretraining stimulus was reduced from 6.2 to 5.0 cm and the training stimuli were reduced from 4.7 to 4.5 cm. This reduction allowed the circular stimuli to fit within the three columns inherent to the test chamber (see Fig. 1a). All of the positions were 13.3 cm from the chamber floor. The middle position was at the midpoint of the iPad and the chamber, and the left and right positions were 15 cm, center to center. The pretraining stimulus was displayed in each of three positions during pretraining but the training stimuli occupied only the left and right positions during simultaneous discrimination training.

4.2.2. Pretraining

Subjects were trained to touch the pretraining stimulus in the left, middle, and right positions on the iPad (see Wolf et al.,



Fig. 4. Data from the first 4-s of trials in Experiments 2 and 3 were used to calculate discrimination ratios for each rat. Data from the last session of Experiment 2, as well as the last session of Phase 1 (Day 3 with correction method) and Phase 2 (Day 9 with non-correction method) of Experiment 3 for all rats. The dotted lines represent animals that received a flashing light after an incorrect response in Experiment 2. The dotted line at .50 represents chance performance.

submitted for publication, for details). After two days of training, rats were then trained with the pretraining stimulus at both the left and right positions, but the position associated with reinforcement was determined pseudo randomly so that it occurred equally often at both sides. Training continued for two days or until any side bias was eliminated.

4.2.3. Phase 1: correction simultaneous discrimination training (Days 1–3)

Each training session consisted of 50 trials. The first trial of every session was a probe trial, in which the trial duration was 60-s and there was no opportunity for reinforcement. On the remaining trials, a trial terminated with a touch to the S+ after the FI-4 schedule or 180 s elapsed from trial onset, whichever came first. Responses to the S- were recorded but had no nominal effects. A 4-s fixed-time ITI separated all trials. The houselight was off during all trials, but remained on during the ITI period.

4.2.4. Phase 2: non-correction simultaneous discrimination training (Days 4–9)

Training was similar to that of Phase 1, with the exception that a response to the S– after the FI-4 terminated the trial with a flashing light, non-reinforcement, and a 16-s timeout for all rats. A programming error resulted in the first trial terminating with a response to the S– for all rats. The data from these trials were discarded from the analysis.

4.3. Results

4.3.1. Phase 1 (Days 1-3)

During the probe trial on the first trial of Day 1, a preference to respond to the S+ was observed for 2 of the 4 rats (binomial tests, ps < .05). The two rats that performed above chance were the same rats that demonstrated discrimination learning in Experiment 2. This same result was found if we conduct the analysis on responding during the first 4 s (i.e., the FI-4 s period) across all of the trials on Day 1 (binomial tests, ps < .05). Fig. 4 displays performance from the last session of Experiment 2, as well as the last sessions of Phase 1 and 2 of Experiment 3. We find evidence of learning in all rats when the analysis was conducted on data from Days 2–3, such that all rats are responding more to the S+(M=179.75, SD = 190.42) than

S - (M = 65.25, SD = 35.11) during the first 4 s of the trials (binomial tests, ps < .05).

4.3.2. *Phase 2 (Days 4–9)*

Given the duration of Phase 1 (3 Days), it would be informative to evaluate performance on Days 5–6 of Phase 2. Performance during the first 4s of trials revealed that 2 of the 4 rats that were responding above chance on Days 2–3 of Phase 1, responded below chance on Days 5–6 (binomial tests, ps > .05). The two rats that performed below chance on Days 5–6 were the same rats that showed no evidence of learning in Experiment 2. As indicated by Fig. 4, performance in these rats did not improve after an additional 3 training sessions (Day 9) with the non-correction method.

Successful performance by rats in Experiment 2 (successive discrimination) led to a preference to respond to the S+ on the very first trial of a simultaneous discrimination in Experiment 3, despite novel positions for the S+ and S-. Interestingly, the two rats that failed to learn in Experiment 2 provided evidence of learning with a correction method during the simultaneous discrimination task. Performance decreased, however, across six subsequent training sessions with the non-correction method. The debate over the benefits of correction versus non-correction methods has a long history in learning, dating back to at least Kalish (1946) and continuing in recent research (e.g., Metcalfe et al., 2009). On the one hand, a correction method allows for each trial to terminate in reinforcement; on the other hand, it also allows for the incorrect response to be followed closely in time by reinforcement thereby creating response chains (Sutherland and Mackintosh, 1971). Our results suggest that rats do benefit from the opportunity to correct their mistakes early in training. The correction procedure allows an incorrect response to the S- to be quickly compared to a correct response to the S+, whereas, a rat with a non-correction method must endure the duration of the ITI before another opportunity to view the S+ and emit a response. Rats already responding correctly to a successive discrimination did not benefit from the opportunity to quickly correct mistakes. Though in need of subsequent validation, these results suggest the use of a correction method early in discrimination training.

One potential concern with our procedure is that a nonrewarded response on a trial could simply become a cue for a rat to respond to the other stimulus. For this reason, our analysis was confined to responses during the FI-4 s portion of the trial. During this time, responses to both the S+ and the S- were non-rewarded and thus cued switching should result in chance performance. Another concern is that the left and right positions occupied by the S+ and Swere separated by 15 cm and an empty middle panel (see Fig. 1a). This distance prevented simultaneous comparison of both stimuli from close distances. This factor likely contributed to the failures of the two rats to learn during non-correction training. In future studies, the stimuli could be placed in adjacent panels, reducing the distance from 15 to 7.5 cm. In our experimental setup, the left, center, and right panels are a consequence of the Med-Associates test chamber being used. A custom-built chamber could have the benefit of a uniform surface but may then lack the flexibility of using modular Med-Associates components (e.g., levers, lights, speakers, etc.).

5. General discussion

The present series of experiments evaluated whether rats could interact with an iPad and discriminate visual stimuli presented on the iPad display. Experiment 1 found that rats will quickly learn to interact with an iPad but fail to learn in a successive non-match to sample task with patterned stimuli. In Experiment 2, two of four rats responded more to a visual stimulus associated with reward than a stimulus associated with a time-out and flashing light in a successive discrimination task. One rat learned with only the flashing light (differential outcome), but both rats improved dramatically with a DRO schedule. In Experiment 3, all four rats were trained in a simultaneous discrimination task. The two rats that learned the discrimination in the second experiment performed well on the very first trial with the S+ and the S– displayed simultaneously in novel locations. Subsequent training with a correction method resulted in a preference for responding to the S+ for all rats. After switching to a non-correction method, performance decreased for two rats. These findings suggest the iPad can be a useful tool in behavioral investigations of visual learning and memory, but also indicate more research is needed to identify the optimal conditions for acquisition of discriminations (e.g., Bussey et al., 2008).

The continued use of touchscreen procedures is encouraged by successful demonstration of simultaneous discrimination and reversals (e.g., Bussey et al., 1997a; Chudasama et al., 2001; Morton et al., 2006); visuospatial conditional discrimination and reversal (e.g., Bussey et al., 1997a; Chudasama et al., 2001; Janisewicz and Baxter, 2003); configural discrimination tasks (Bussey et al., 1998); nonspatial nonmatch-to-sample (e.g., Bussey et al., 1994); and autoshaping (e.g., Dalley et al., 2005; Bussey et al., 1997a; Parkinson et al., 1999, 2002) in rats and mice using a touchscreen. These types of tasks previously required fabrication of a custom apparatus or costly touchscreen packages, but the current experiments demonstrate that the iPad presents a relatively simple alternative, utilizing equipment that many behavioral scientists are using or have access to. One additional benefit of the iPad over infrared touchscreens includes the opportunity for multi-touch detection. This feature could be used to require cooperation or competition between rats working on the same display or in observational learning paradigms.

A synergistic relationship exists between new technologies and scientific inquiry. Better technology allows the scientist to accumulate more evidence regarding existing questions but also expands the potential for new questions. The result of this relationship can be seen within the research career of Dr. Anthony Wright. Wright and colleagues pioneered the use of touchscreen technology for pigeons (Wright et al., 1988) and primates (Bhatt and Wright, 1992); concluding that touchscreen-equipped monitors could enhance stimulus control and flexibility, while maintaining precise response detection. In subsequent studies, Wright and colleagues used touchscreens to conduct comparative investigations of concept learning, memory effects in serial list learning, and working memory capacity. The present experiments corroborate Dr. Anthony Wright's legacy of incorporating new technology and examining the influence of experimental parameters in the study of learning and memory.

References

- Allan, R.W., 1992. Technologies to reliably transduce the topographical details of pigeons' pecks. Behav. Res. Methods Instrum. Comput. 24, 150–156.
- Bhatt, R.S., Wright, A.A., 1992. Concept learning by monkeys with video picture images and a touch screen. J. Exp. Anal. Behav. 57, 219–225.
- Blough, D.S., 1986. Odd-item search by pigeons: method, instrumentation, and uses. Behav. Res. Methods Instrum. Comput. 18, 413–419.
- Bussey, T.J., Dias, R., Amin, E., Muir, J.L., Aggleton, J.P., 2001. Perirhinal cortex and place-object conditional learning in the rat. Behav. Neurosci. 115, 776–785.
- Bussey, T.J., Everitt, B.J., Robbins, T.W., 1997a. Dissociable effects of cingulate and medial frontal cortex lesions on stimulus-reward learning using a novel Pavlovian autoshaping procedure for the rat: implications for the neurobiology of emotion. Behav. Neurosci. 111, 908–919.
- Bussey, T.J., Muir, J.L., Everitt, B.J., Robbins, T.W., 1997b. Triple dissociation of anterior cingulate, posterior cingulate, and medial frontal cortices on visual discrimination tasks using a touchscreen testing procedure for the rat. Behav. Neurosci. 111, 920–936.
- Bussey, T.J., Muir, J.L., Robbins, T.W., 1994. A novel automated touchscreen procedure for assessing learning in the rat using computer graphic stimuli. Neurosci. Res. Commun. 15, 103–110.

- Bussey, T.J., Padain, T.L., Skillings, E.A., Winters, B.D., Morton, A.J., Saksida, L.M., 2008. The touchscreen cognitive testing method for rodents: how to get the best out of your rat. Learn Mem. 15 (7), 516–523.
- Bussey, T.J., Warburton, E.C., Aggleton, J.P., Muir, J.L., 1998. Fornix lesions can facilitate acquisition of the transverse patterning task: a challenge for configural theories of hippocampal function. J. Neurosci. 18, 1622–1631.
- Chudasama, Y., Bussey, T.J., Muir, J.L., 2001. Effects of selective thalamic and prelimbic cortex lesions on two types of visual discrimination and reversal learning. Eur. J. Neurosci. 14, 1009–1020.
- Cook, R.G., Geller, A.I., Zhang, G.R., Gowda, R., 2004. Touchscreen-enhanced visual learning in rats. Behav. Res. Methods Instrum. Comput. 36, 101–106.
- Dalley, J.W., Lääne, K., Theobald, D.E., Armstrong, H.C., Corlett, P.R., Chudasama, Y., Robbins, T.W., 2005. Time-limited modulation of appetitive Pavlovian memory by D1 and NMDA receptors in the nucleus accumbens. Proc. Natl. Acad. Sci. U.S.A. 102, 6189–6194.
- Delamater, A.R., Kranjec, A., Fein, M.I., 2010. Differential outcome effects in Pavlovian biconditional and ambiguous occasion setting tasks. J. Exp. Psychol. Anim. Behav. Process. 36, 471–481.
- Dudchenko, P.A., 2004. An overview of the tasks used to test working memory in rodents. Neurosci. Biobehav. Rev. 28, 699–709.
- Elsmore, T.F., Parkinson, J.K., Mellgren, R.L., 1989. Video touchscreen stimulusresponse surface for use with primates. Bull. Psychon. Soc. 27 (1), 60–63.
- Frank, A.J., Wasserman, E.A., 2005. Associative symmetry in the pigeon after successive matching-to-sample training. J. Exp. Anal. Behav. 84 (2), 147–165.
- Hall, G., 1991. Perceptual and Associative Learning. Clarendon Press, Oxford.
- Huguenin, N.H., 2000. Reducing overselective attention to compound visual cues with extended training in adolescents with severe mental retardation. Res. Dev. Disabil. 21, 93–113.
- Hunter, W.S., 1913. The delayed reaction in animals and children. Behav. Monogr. 2, 1–86.
- Janisewicz, A.M., Baxter, M.G., 2003. Transfer effects and conditional learning in rats with selective lesions of medial septal/diagonal band cholinergic neurons. Behav. Neurosci. 117, 1342–1352.
- Kalish, D., 1946. The non-correction method and the delayed response problem of Blodgett and McCutchan. J. Comp. Psychol. 39, 91–108.
- Kelly, R., Spetch, M.L., Grant, D.S., 1999. Influence of nonmemorial factors on manifestation of short-sample biases in choice and successive matching-to-duration tasks with pigeons. J. Exp. Psychol. Anim. Behav. Process. 25 (3), 297–307.
- Lashley, K.S., 1938. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detailed vision. J. Gen. Psychol. 18, 123–193.
- Lieberman, D.A., Davidson, F.H., Thomas, G.V., 1985. Marking in pigeons: the role of memory in delayed reinforcement. J. Exp. Psychol. Anim. Behav. Process. 11, 611–624.
- Lieberman, D.A., McIntosh, D.C., Thomas, G.V., 1979. Learning when reward is delayed: a marking hypothesis. J. Exp. Psychol. Anim. Behav. Process. 5, 224–242.
- Markham, M.R., Butt, A.E., Dougher, M.J., 1996. A computer touch-screen apparatus for training visual discriminations in rats. J. Exp. Anal. Behav. 65, 173–182.
- Metcalfe, J., Kornell, N., Finn, B., 2009. Delayed versus immediate feedback in children's and adults' vocabulary learning. Mem. Cognit. 37 (8), 1077–1087.

- Minini, L., Jeffery, K.J., 2006. Do rats use shape to solve shape discriminations? Learn. Mem. 13, 287–297.
- Morton, A.J., Skillings, E., Bussey, T.J., Saksida, L.M., 2006. Measuring cognitive deficits in disabled mice using an automated interactive touchscreen system. Nat. Methods 3, 767.
- Overmier, J.B., Linwick, D., 2001. Conditional choice: unique outcomes establish expectancies that mediate choice behavior. Integr. Physiol. Behav. Sci. 36, 173–181.
- Parkinson, J.A., Dalley, J.W., Cardinal, R.N., Bamford, A., Fehnert, B., Lachenal, G., Rudarakanchana, N., Halkerston, K.M., Robbins, T.W., Everitt, B.J., 2002. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. Behav. Brain Res. 137, 149–163.
- Parkinson, J.A., Olmstead, M.C., Burns, L.H., Robbins, T.W., Everitt, B.J., 1999. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. J. Neurosci. 19, 2401–2411.
- Pisacreta, R., Rilling, M., 1987. Infrared touch technology as a response detector in animal research. Behav. Res. Methods Instrum. Comput. 19, 389–396.
- Pontecorvo, M.J., 1983. Effects of proactive interference on rats' continuous nonmatching-to-sample performance. Anim. Learn. Behav. 11, 356–366.
- Pontecorvo, M.J., Sahgal, A., Steckler, T., 1996. Further developments in the measurement of working memory in rodents. Cogn. Brain Res. 3, 205–213.
- Prusky, G.T., Harker, K.T., Douglas, R.M., Whishaw, I.Q., 2002. Variation in visual acuity within pigmented, and between pigmented and albino rat strains. Behav. Brain Res. 136, 339–348.
- Sahgal, A., Steckler, T., 1994. Touch Windows and operant behavior in rats. J. Neurosci. Methods 55, 59–64.
- Schmidtke, K.A., Katz, J.S., Wright, A.A., 2010. Differential outcomes facilitate same/different concept learning. Anim. Cogn. 13, 583–589.
- Simpson, E.L., Gaffan, E.A., 1999. Scene and object vision in rats. Q. J. Exp. Psychol. 52B, 1–29.
- Sutherland, N.S., 1961. Shape Discrimination by Animals. Experimental Psychology Society. Monographs, Heffer and Sons, Cambridge, UK.
- Sutherland, N.S., Mackintosh, N.J., 1971. Mechanisms of Animal Discrimination Learning. New York, Academic Press.
- Urcuioli, P.J., Kasprow, W.J., 1988. Long-delay learning in the T-maze: effects of marking and delay-interval location. Learn. Motiv. 19, 66–86.
- Williams, B.A., 1999. Blocking the response-reinforcer association: theoretical implications. Psychon. Bull. Rev. 6, 618–623.
- Wolf, J.E., Urbano, K., Ruprecht, C.M., Leising, K.J., Need to train your rat? There is an app for that: a touchscreen-based behavioral evaluation system. Behav. Res. Methods, submitted for publication.
- Wright, A.A., 2010. Functional relationships for determining similarities and differences in comparative cognition. Behav. Process. 85, 246–251.
- 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. Anim. Learn. Behav. 16, 436–444.
- Wright, A.A., Delius, J.D., 1994. Scratch and match: pigeons learn matching and oddity with gravel stimuli. J. Exp. Psychol. Anim. Behav. Process. 20, 108–112.

Contents lists available at SciVerse ScienceDirect







journal homepage: www.elsevier.com/locate/behavproc

Role of the hippocampus in mediating interference as measured by pattern separation processes

Raymond P. Kesner*

University of Utah, Salt Lake City, UT, USA

ARTICLE INFO

Article history: Received 12 July 2012 Received in revised form 3 September 2012 Accepted 7 September 2012

Keywords: Hippocampus Dentate gyrus CA3 CA1 Spatial, temporal, and odor pattern separation

ABSTRACT

In order to understand the neural mechanism associated with specific forms of interference, this manuscript concentrates on the role of the dorsal and ventral dentate gyrus subregions of the hippocampus in rats. The computational modelers have suggested that the dentate gyrus can provide a neural mechanism that can operate to reduce interference between highly processed similar spatial, contextual or odor inputs to generate pattern separation functions. Pattern separation which is defined as a process to remove redundancy from similar inputs so that events can be separated from each other and interference can be reduced, and in addition can produce a more orthogonal, sparse, and categorized set of outputs. It appears that the anatomical organization of the hippocampus may provide the answer for the importance of interference in mnemonic processing of information. Therefore, in the first part of this paper an anatomical description of the inputs and outputs of the hippocampus as well as its intrinsic circuit is provided. This is followed by the presentation of data to support the role of the dorsal DG in supporting spatial pattern separation, dorsal CA1 in supporting temporal pattern separation for odors.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Tony Wright has had a profound influence on many aspects of animal cognition. He has emphasized the importance of interference in processing information in working memory, list learning and concept formation, see a review by Wright (2007). Is there a neurobiological mechanism that can determine the level of interference that affects the level of performance in a variety of cognitive tasks? Early computational models and reviews of the literature by O'Reilly and McClelland (1994), Shapiro and Olton (1994) and Rolls (1989) have suggested that the hippocampus and especially the dentate gyrus (DG) could support pattern separation which is defined as a process to remove redundancy from similar inputs so that events can be separated from each other and interference can be reduced, and in addition can produce a more orthogonal, sparse, and categorized set of outputs. The cells of the DG are suggested to act as a competitive learning network with Hebb-like modifiability to reduce redundancy and produce sparse, orthogonal outputs. It appears that the anatomical organization of the hippocampus may provide the answer for the importance of interference in mnemonic processing of information. Therefore, in the first part of this paper an anatomical description of the inputs and outputs of the hippocampus as well as its intrinsic circuit is provided. This is followed by the presentation of data to support the role of the dorsal DG in supporting spatial pattern separation, dorsal CA1 in supporting temporal pattern separation for spatial locations and visual objects, ventral DG in supporting odor pattern separation, and ventral CA1 in supporting temporal pattern separation for odors.

2. Anatomical description of the hippocampus

A schematic representation of the intrinsic circuit of the hippocampus and its entorhinal connections are shown in Fig. 1. The main input into the hippocampal system is from entorhinal cortex which receives inputs from multiple cortical regions and all sensory modalities. The cortical inputs that terminate on the superficial layers (I, II, and III) of the entorhinal cortex comprise the primary inputs to the hippocampus (Witter, 1993). In the rat, the cortical inputs to the superficial layers of entorhinal cortex originate in the olfactory domain of the telencephalon, perirhinal cortex, and pre- and parasubiculum. As noted by Witter (1993), entorhinal cortex then projects directly to three hippocampal structures. Cells in layer II of the entorhinal cortex project primarily to the dentate gyrus and also to CA3/2. The projections

^{*} Correspondence address: Department of Psychology, University of Utah, 380 South 1530 East, Rm 502, Salt Lake City, UT 84112, USA. Tel.: +1 801 581 7430; fax: +1 801 581 5841.

E-mail address: ray.kesner@psych.utah.edu

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.09.018



Fig. 1. Schematic illustration of the principal connections from the lateral entorhinal cortex (LEC) and medial entorhinal cortex (MEC) to the apical dendrites of the hippocampal pyramidal neurons in CA1 and CA3a,b,c and granule cells in the dentate gyrus (DG). II and III indicate layers of the entorhinal cortex. Perforant path (PP), anterior commissure (AC) and subiculum (SC).

that terminate in the CA1 region originate in layer III of the entorhinal cortex. The primary projection of the entorhinal cortex is to the dentate gyrus. The connections between entorhinal cortex, dentate, and CA3 are generally reported to be feed-forward (Ishizuka et al., 1990; Witter, 1993). The dentate gyrus granular neurons project to CA3 pyramidal neurons via mossy fiber projections. The neurons that comprise CA3, in turn, project to CA1 neurons via the Schaffer collaterals. Recurrent collateral connections exist within both the dentate gyrus and CA3 that serve to interconnect neurons within these respective regions. The most prominent anatomical feature of the CA3 subregion is that there are extensive interconnections among the principal cells via a recurrent collateral fiber system (Amaral and Witter, 1995). CA3 also receives converging inputs from multiple input pathways; for example, perforant path inputs from the medial and lateral entorhinal cortex, mossy fiber inputs from the dentate gyrus, and its own outputs feed back as inputs via the recurrent collaterals (Amaral and Witter, 1995). The primary output from the hippocampus to neocortex originates in CA1 and projects to subiculum, entorhinal cortex, and parahippocampal structures (Witter, 1993), but there are also projections that originate in CA3. Many researchers have reported that CA3 projects to the lateral and medial septal nuclei as well as to the vertical limb of the diagonal band of Broca (Gaykema et al., 1991; Amaral and Witter, 1995; Risold and Swanson, 1997). The medial septum and vertical limb of the diagonal band of Broca, in turn, provides cholinergic and GABAergic inputs into the hippocampus (Amaral and Witter, 1995). It has been shown that the CA3 region can be divided into a CA3a, b, and c subareas (Li et al., 1994; Lorente de Nó, 1934). Based on the research of Li et al. (1994) and Buckmaster and Schwartzkroin (1994), it has been proposed that mossy cells receive excitatory inputs from granule cells and CA3c pyramidal cells and integrate the inputs from granule cells and CA3c pyramidal cells, which, in turn, via excitatory recurrent axonal projections activate many distal granule cells. Thus, CA3c may have a back-projection output that can influence the DG granule cells (Scharfman, 2007). Most of the synaptic connections embedded in those different pathways in the different subregions of the hippocampus are modifiable in their strength (Marr, 1971; Treves and Rolls, 1994). It should be noted that lesions will usually involve the dorsal hippocampus unless otherwise stated and CA3 lesions will include CA3a, b unless otherwise stated. These anatomical and physiological characteristics inspired many theoretical models to assign specific cognitive processes to the different subregions (DG, CA3, and CA1) of the hippocampus (Marr, 1971; O'Reilly and McClelland, 1994; Treves and Rolls, 1994; Rolls, 1996; Samsonovich and McNaughton, 1997; Lisman, 1999; Mizumori et al., 1999; Hasselmo and Wyble, 1997; Hasselmo et al., 2002; Kesner et al., 2004; Rolls and Kesner, 2006).

3. Spatial pattern separation

3.1. Computational models

The determination of a spatial pattern separation process has been developed extensively by computational models of the subregions of the hippocampus with a special emphasis on the DG. Based on the empirical findings that all sensory inputs are processed by the DG subregion of the hippocampus (Aggleton et al., 1986; Jackson-Smith et al., 1993; Kesner et al., 1993; Mumby et al., 1992; Otto and Eichenbaum, 1992), it has been suggested that a possible role for the hippocampus might be to provide for sensory markers to demarcate a spatial location, so that the hippocampus can more efficiently mediate spatial information. It is thus possible that one of the main process functions of the hippocampus is to encode and separate spatial events from each other. This would ensure that new highly processed sensory information is organized within the hippocampus and enhances the possibility of remembering and temporarily storing one place as separate from another place. It is assumed that this is accomplished via pattern separation of event information, so that spatial events can be separated from each other and spatial interference is reduced. This process is akin to the idea that the hippocampus is involved in orthogonalization of sensory input information (Rolls, 1989), in representational differentiation (Myers et al., 1995), and indirectly in the utilization of relationships (Cohen and Eichenbaum, in press).

Rolls' (1996) model based in part on Marr's ideas (Marr, 1971) proposes that pattern separation is facilitated by sparse connections in the mossy-fiber system, which connects DG granular cells to CA3 pyramidal neurons in combination with a competitive inhibition network associated with the granule cells of the dentate gyrus. Separation of patterns is accomplished based on the low probability that any two CA3 neurons will receive mossy fiber input synapses from a similar subset of DG cells. Mossy fiber inputs to CA3 from DG are suggested to be essential during learning and may influence which CA3 neurons fire based on the distributed activity within the DG. Cells of the DG are suggested to act as a competitive learning network with Hebb-like modifiability to reduce redundancy and produce sparse, orthogonal outputs. O'Reilly and McClelland (1996) and Shapiro and Olton (1994) also suggested that the mossy fiber connections between the DG and CA3 may support pattern separation.

3.2. Behavioral evidence in rodents and humans

To examine the contribution of the DG to spatial pattern separation, Gilbert et al. (2001) tested rats with DG lesions using a paradigm which measured short-term memory for spatial location information as a function of spatial similarity between spatial locations. Specifically, the study was designed to examine the role of the DG subregion in discriminating spatial locations when rats were required to remember a spatial location based on distal environmental cues and to differentiate between the to-be-remembered location and a distractor location with different degrees of similarity or overlap among the distal cues.

Animals were tested using a cheeseboard maze apparatus (the cheese board is similar to a dry land water maze with 177 circular, recessed holes on a 119 cm diameter board) on a delayed-match-tosample for a spatial location task. Animals were trained to displace an object which was randomly positioned to cover a baited food well in 1 of 15 locations along a row of food wells. Following a short delay, the animals were required to choose between objects which were identical to the sample phase object: one object was in the same location as the sample phase object and the second object was in a different location along the row of food wells. Rats were rewarded for displacing the object in the same spatial location as the sample phase object (correct choice), but they received no reward for displacing the foil object (incorrect choice). Five spatial separations, from 15 cm to 105 cm, were used to separate the correct object and the foil object during the choice phase. Rats with DG lesions were significantly impaired at short spatial separations; however, during the choice phase performance of DG-lesioned animals increased as a function of greater spatial separation between the correct and foil objects. The performance of rats with DG lesions matched control rats at the largest spatial separation. The graded nature of the impairment and the significant linear improvement in performance as a function of increased separation illustrate a deficit in pattern separation. Based on these results, it was concluded that lesions of the DG decrease the efficiency of spatial pattern separation, which results in impairments on trials with increased spatial proximity and increased spatial similarity among working memory representations. Holden et al. (2012) used an analogous task to that used for rats (Gilbert et al., 2001) to test young participants compared to aged participants who are likely to have DG dysfunction (see Small et al., 2011). They report that aged participants that do not perform well on standard memory tests are impaired in displaying a pattern separation function. One limitation of the dot task is that it does not assess the ability to separate spatial patterns in the real world. In order to assess real world spatial pattern separation, hypoxic subjects with hippocampal damage and matched normal controls were administered a geographical spatial distance task (cities on a map; Hopkins and Kesner, 1993). The subjects were shown 8 cities on a map of New Brunswick one at a time for 5 s each. Subjects were instructed to remember the city and its spatial location on the map. In the test phase the subjects were presented with the names of two cities that occurred in the study phase and were asked which of the cities was located further to the east (on separate trials, subjects were asked which city occurred further north, south, or west). There were two trials for each compass direction. Spatial distances of 0, 2, 4, and 6 as measured by the number of cities in the study phase that were geographically situated between the two test cities were measured. There were 8 trials for each distance. The hypoxic subjects were impaired for all spatial distances for spatial geographical information compared to control subjects who displayed a pattern separation function for distance (Hopkins and Kesner, 1993). Thus, the DG may function to encode and to separate locations in space to produce spatial pattern separation. Such spatial pattern separation ensures that new highly processed sensory information is organized within the hippocampus, which in turn enhances the possibility of encoding and temporarily remembering one spatial location as separate from another.

3.3. Distance representations in rodents and humans

Based on the observations that cells in CA3 and CA1 regions respond to changes in metric aspects of the environment (Jeffery and Anderson, 2003; O'Keefe and Burgess, 1996), one can ask whether these different features of the spatial environment are processed via the DG and then subsequently transferred to the CA3 subregion or if these features are communicated via the direct perforant path projection to the CA3 subregion. In both cases, information may then be transferred to the CA1 subregion.

To answer this question, Goodrich-Hunsaker et al. (2005) examined the contributions of the DG to memory for metric spatial relationships. Using a modified version of an exploratory paradigm developed by Poucet (1989), rats with DG, CA3, and CA1 lesions as well as controls, were tested on tasks involving a metric spatial manipulation. In this task, a rat was allowed to explore two different visual objects separated by a specific distance on a cheeseboard maze. On the initial presentation of the objects, the rat explored each object. However, across subsequent presentations of the objects in the same spatial locations, the rat habituated and eventually spent less time exploring the objects. Once the rat had habituated to the objects was manipulated so that the two objects were either closer together or farther apart. The time the rat spent exploring each moved object was recorded.

The results showed that rats with DG lesions spent significantly less time exploring the two objects that were displaced relative to controls, indicating that DG lesions impair the detection of metric distance changes. Rats with CA3 or CA1 lesions displayed mild impairments relative to controls, providing empirical validation for the role of DG in spatial pattern separation and support the predictions of computational models (Rolls, 1996; Rolls and Kesner, 2006). Stark et al. (2010) used an analogous task to that used for rats (Goodrich-Hunsaker et al., 2005) to measure spatial pattern separation based on distance, and in this case angle as well, to test young and healthy aging humans. Even though there are some individual differences, they report an impairment in spatial pattern separation. Also, Baumann et al. (2012) reported activation of the posterior hippocampus in spatial pattern separation using the task used by Goodrich-Hunsaker et al. (2005).

3.4. Picture representations in rodents and humans

In other studies with humans a modified continuous recognition task was used. Using high resolution fMRI in this task revealed in the first study with young participants that the hippocampus distinguished between correctly identified true stimulus repetitions, correctly rejected presentations of similar lure stimuli, and false alarm lures (Kirwan and Stark, 2007). In a subsequent study it was shown that in aged compared to young participants that the DG/CA3 subregions of the hippocampus played an important role in deficits found in aged participants (Yassa et al., 2010). For a review of the human pattern separation data see Yassa and Stark (2011). It is important to note that in rats the perirhinal cortex supports pattern separation for visual objects and visual features (Bussey et al., 2002; Gilbert and Kesner, 2003). One possible explanation might be that perhaps humans tend to add verbal labels to each visual object, which would then be consistent with the role that the hippocampus plays in processing linguistic information.

3.5. Spatial pattern separation modulation of CA3 and CA1 functions

Because of the feed-forward connections of the DG with CA3 and CA1, one might expect that a disruption of pattern separation may also influence paired associate learning mediated by CA3 and spatial learning tasks mediated by CA1 and CA3 as well as temporal memory retrieval for places mediated by CA1. As an example, disruption of DG and mossy fiber inputs into CA3 do not produce a disruption in the acquisition of an object-place paired associate task (Gilbert and Kesner, 2003) unless the stimuli are close together, implying that the DG contribution is important particularly when pattern separation is needed (Gilbert and Kesner, unpublished observations). The implication is that sufficient input for objectplace learning can be introduced into the CA3 system (which is required for this object-place learning) by the perforant path inputs provided that spatial pattern separation is not at a premium. Also, based on the McDonald and White (1995) place preference procedure in an eight-arm maze, in which food is placed at the end of one arm and no food is placed at the end of another arm, it was shown that rats with dorsal DG lesions and control lesions acquired the spatial discrimination for separate locations at similar rates. However, for the adjacent condition, dorsal DG lesioned rats required significantly more trials to reach the learning criterion than controls (Morris et al., 2011). In a final experiment, rats with DG lesions were tested in a previously mentioned temporal ordering of spatial location task (Hunsaker and Kesner, 2008), but in this case with a small distance of 54 or 108 cm at a 30 min delay between the study phase and the test phase consisting of the A vs. C preference test. The results indicate that animals with DG lesions showed intact preference for A similar to controls with 108 cm spatial separation, but had a deficit in preferring C with a 54 cm spatial separation, suggesting that the DG is involved in processing spatial information when there is a difficulty associated with high levels of spatial interference resulting in a spatial pattern separation problem. Thus, spatial pattern separation may play an important role in the acquisition of new spatial information and there is a good possibility that the DG may have been the subregion responsible for the impairments in the various tasks described above.

3.6. Role of neurogenesis in the DG on spatial pattern separation

Based on the observation that neurogenesis occurs in the DG and that new DG granule cells can be formed across time, it has been proposed that the DG mediates a spatial pattern separation mechanism as well as generates patterns of episodic memories within remote memory (Aimone et al., 2006). Thus far, it has been shown in mice that disruption of neurogenesis using low-dose x-irradiation was sufficient to produce a loss of newly born DG cells. Further testing indicated impairments in spatial learning in a delayed non-matching-to-place task in the radial arm maze. Specifically, impairment occurred for arms which were presented with little separation, but no deficit was observed when the arms were presented farther apart, suggesting a spatial pattern separation deficit. Also, the disruption of neurogenesis using lentivirus expression of a dominant Wnt protein produced a loss of newly born DG cells, as well, and was observed in an associative object-in-place task with different spatial separations as a function of the degree of separation, again suggesting a spatial pattern separation deficit (Clelland et al., 2009). In a more recent study (Kesner and Fanselow, unpublished observations) it can be shown that DNA methyltransferase 1-c knockout mice are impaired relative to controls in the Goodrich-Hunsaker et al. (2005) spatial pattern separation task. These data suggest that neurogenesis in the DG may contribute to the operation of spatial pattern separation. Thus, spatial pattern separation may play an important role in the acquisition of new spatial information and there is a good possibility that the DG may be the subregion responsible for the impairments in the various tasks described above.

3.7. Spatial context and pattern separation

Do other subregions of the hippocampus engage in spatial pattern separation? It appears that the CA3 region may also engage spatial pattern separation processes, but in this case not for a metric spatial representation, but for spatial representation of the geometry of the environment. This idea is supported by the finding of Tanila (1999) who showed that CA3c place cells were able to maintain distinct representations of two visually identical environments, and selectively reactivate either one of the representation patterns depending on the experience of the rat. Also, Leutgeb et al. (2007) recently showed that when rats experienced a completely different environment, CA3c place cells developed orthogonal representations of those different environments by changing their firing rates between the two environments, whereas CA1 place cells maintained similar responses. To further test the role of CA3 in mediating pattern separation, an experiment was conducted to determine whether the DG or CA3 regions cooperate to perform spatial pattern separation operations for specific spatial locations as well as the spatial geometry of the environment or whether the DG performs spatial pattern separation on the basis of specific locations in space and the CA3 performs spatial pattern separation on the basis of the geometry of the environment (Hunsaker et al., 2008b). Rats with lesions of DG and CA3a,b were given the opportunity to explore a white or black circular or square box of the same size as reported by Leutgeb et al. (2007) and, in addition, in the box there were two objects spaced 68 cm apart. After habituation to the box and the objects, the rats received one of two transfer tests. In the first test the objects were changed to a 38 cm distance, but the box shape (geometry of the environment) remained the same. In the second test the box shape (geometric environment) was changed, but the distance between the objects remained the same. The efficacy of the transfer test in terms of re-exploration of the metric change is based on a comparison between the level of object exploration during the transfer session versus object exploration during the last session of habituation. Similarly, the efficacy of the transfer test in terms of re-exploration of the geometry of the environment is based on the number of grid crossings (activity level) and rearings during the transfer session versus the number of grid crossings and rearings during the last session of habituation. The results indicate that lesions of the DG, but not CA3a,b, disrupt both the detection of metric changes in the spatial location of objects and changes in a geometrical environment. Thus far, these data are consistent with the prediction of the Rolls computational model that the DG is the critical substrate for spatial pattern separation. These data are not consistent with (Tanila, 1999; Leutgeb et al., 2007) findings of a pattern separation function for geometrical environments. It has been shown that the CA3 region can be divided into a CA3a,b, and c subareas (Li et al., 1994; Lorente de Nó, 1934). Most of the recorded cells that respond to different environments reported by Tanila (1999) and Leutgeb et al. (2007) were based on electrode placements in the CA3c area. The lesion data were based on lesions within CA3a/b, but not CA3c. Based on the research of Li et al. (1994) and Buckmaster and Schwartzkroin (1994), it has been proposed that mossy cells receive excitatory inputs from granule cells and CA3c pyramidal cells and integrate the inputs from granule cells and CA3c pyramidal cells, which, in turn, via excitatory recurrent axonal projections activate many distal granule cells. Such a circuit could integrate spatial location information and form representations of geometrical environments. Additional experiments with CA3c lesions in contrast to the CA3a/b lesions were carried out (Hunsaker et al., 2008a). The results indicated that dorsal CA3c lesions only disrupted pattern separation processes when the animal was required to detect a metric change in object location, but there was no apparent effect during the environmental change task. It must be noted, however, that dorsal CA3c lesions never caused effects as dramatic as those caused by dorsal dentate gyrus lesions. One interpretation may be that the dentate gyrus selectively recruits CA3c to assist in the metric detection and not the detection of the overall environmental change. The present experiment provides behavioral evidence that dorsal CA3c and the dorsal dentate gyrus may interact for spatial information processing. This effect was only seen during the condition in which the animal is required to detect a discrete metric change in object location, a task that has been shown to be particularly sensitive to dentate gyrus damage (Goodrich-Hunsaker et al., 2005). In a new experiment based on novelty detection for a change in the color (black, dark gray, light gray and white) of the box (context), it was possible to show that the DG plays an important role in mediating pattern separation for changes in color (context) (Kesner and Musso, unpublished observations).

In summary, although the present behavioral data do not show any effects of dorsal CA3a,b lesions, the dorsal dentate gyrus effect is clear. Additionally, the dorsal CA3c lesion data suggest that there is a circuit involving dorsal CA3c and the dorsal dentate gyrus that is perhaps important for pre-processing spatial information prior to dorsal CA3a,b processing stages.

It is important to note that other forms of pattern separation do not involve the DG subregion of the hippocampus. For example, temporal pattern separation is mediated by the CA1, but not the DG (Gilbert et al., 2001). Furthermore, hippocampal lesions including DG do not produce a deficit for pattern separation of reward values, visual objects, or motor responses (Gilbert and Kesner, 2002, 2003; Saksida et al., 2006). Instead, the perirhinal cortex subserves pattern separation for visual objects (Bussey et al., 2002; Gilbert and Kesner, 2003; Norman and Eacott, 2004), the amygdala subserves pattern separation for reward value (Gilbert and Kesner, 2002) and the caudate nucleus subserves pattern separation for motor responses (Kesner and Gilbert, 2006).

4. Temporal pattern separation for spatial locations and visual objects

There are data to support the existence of memory for order information, but it is not always clearly demonstrated whether memory for a particular sequence has been learned and can be accurately recalled. Estes (1986) summarized data demonstrating that, in human memory, there are fewer errors for distinguishing items (by specifying the order in which they occurred) that are far apart in a sequence than those that are temporally adjacent. Other studies have also shown that order judgments improve as the number of items in a sequence between the test items increases (Banks, 1978; Chiba et al., 1994; Madsen and Kesner, 1995). This phenomenon is referred to as a temporal distance effect [sometimes referred to as a temporal pattern separation effect (Kesner et al., 2004)]. The temporal distance effect is assumed to occur because there is more interference for temporally proximal events than for temporally distant events.

Based on these findings, Gilbert et al. (2001) tested memory for the temporal order of items in a one-trial sequence learning paradigm in rodents. In the task, each rat was given one daily trial consisting of a sample phase followed by a choice phase. During the sample phase, the animal visited each arm of an 8-arm radial maze once in a randomly predetermined order and was given a reward at the end of each arm. The choice phase began immediately following the presentation of the final arm in the sequence. In the choice phase, two arms were opened simultaneously and the animal was allowed to choose between the arms. To obtain a food reward, the animal had to enter the arm that occurred earlier in the sequence that it had just followed. Temporal separations of 0, 2, 4, and 6 were randomly selected for each choice phase. These values represented the number of arms in the sample phase that intervened between the arms that were to be used in the test phase. After reaching criterion, rats received CA1 or control lesions. Following surgery, control rats matched their preoperative performance across all temporal separations. In contrast, rats with CA1 lesions performed at chance across 0, 2, or 4 temporal separations and a little better than chance in the case of a separation of 6 items. The results suggest that the CA1 subregion is involved in memory for spatial location as a function of temporal separation of spatial locations. Thus, lesions of the CA1 decrease efficiency in temporal pattern separation. CA1 lesioned rats cannot separate events across time, perhaps due to an inability to inhibit interference that may be associated with sequentially occurring events. The increase in temporal interference impairs the rat's ability to remember the order of specific events. Tolentino et al. (2012) used an analogous task to that used for rats (Gilbert et al., 2001) to test young compared to non-demented older participants in a spatial temporal pattern separation task and report temporal pattern separation problems for the older participants. In another spatial location task patients with a hypoxic condition and hippocampal damage are impaired in displaying a temporal pattern separation function (Hopkins et al., 1995).

In a more recent experiment using a paradigm described by Hannesson et al. (2004), it can be shown that temporal order information for spatial location is impaired only for CA1 (Hunsaker et al., 2008b). Thus, it is suggested that the CA1 hippocampal subregion serves as a critical substrate in supporting temporal pattern separation for spatial information.

It has been suggested that the perirhinal cortex and CA1 subregion of the hippocampus play an important role in supporting temporal processing of visual object information (Hoge and Kesner, 2007; Hunsaker et al., 2008b). In humans it can be shown that a temporal pattern separation process can be observed in hypoxic patients in a temporal order memory test for abstract figures (Hopkins et al., 1995), suggesting that the hippocampus may also play a role in temporal pattern separation for visual stimuli at least in humans.

5. Temporal pattern separation for odors

Does the hippocampus support temporal pattern separation processes for sensory-perceptual information other than space and visual objects? To answer this question memory for the temporal order for a sequence of odors was assessed in rats based on a varied sequence of five odors, using a similar paradigm described for sequences of spatial locations. Kesner et al. (2002) found that rats with hippocampal lesions were impaired relative to control animals for memory for all temporal distances between the odors, despite an intact ability to discriminate between the odors. Fortin et al. (2002) reported similar results with fimbria fornix lesions. In a further subregional analysis, rats with dorsal CA1 lesions showed a mild impairment, but rats with ventral CA1 lesions showed a severe impairment in memory for the temporal distance for odors (Kesner et al., 2010). Thus, the CA1 appears to be involved in separating events in time for spatial and non-spatial information, so that one event can be remembered distinct from another event; however, the dorsal CA1 might play a more important role than the ventral CA1 for spatial information (Chiba et al., 1992), and conversely the ventral CA1 might play a more important role than the dorsal CA1 for odor information. The mechanism that could subserve the above mentioned findings is based on the memory question that asks which of two items occurred earlier in the list. To implement this type of memory, some temporally decaying memory trace or temporally increasing memory trace via a consolidation process might provide a model (Marshuetz, 2005); in such a model, temporally adjacent items would have memory traces of more similar strength and would be harder to discriminate than the strengths of the memory traces of more temporally distant items.

6. Odor pattern separation

Working memory and pattern separation for odor information was assessed in rats using a matching-to-sample for odors paradigm. The odor set consisted of a five aliphatic acids with unbranched carbon chains that varied from two to six carbons in length. Each trial consisted of a sample phase followed by a choice phase. During the sample phase, rats would receive one of five different odors. Fifteen sec later, during the choice phase, one of the previous odors was presented simultaneously side by side with a different odor that was based on the number of aliphatic acids that varied in the carbon chains from two to six carbons in length and rats were allowed to choose between the two odors. The rule to be learned in order to receive a food reward was to always choose the odor that occurred during the study phase. Odor separations of 1, 2, 3 or 4 were selected for each choice phase and represented the carbon chain difference between the study phase odor and the test phase odor. Once an animal reached a criterion of 80-90% correct across all temporal separations based on the last 16 trials, rats received a control or ventral dentate gyrus lesion and were retested on the task. On postoperative trials, there were no deficits at 15s delay for either the controls or the ventral dentate gyrus lesioned rats. However, when the delay was increased to 60 s rats with ventral DG lesions were significantly impaired at short spatial separations between the correct and foil objects. The performance of rats with ventral DG lesions matched control rats at the largest odor based separation. The graded nature of the impairment and the significant linear improvement in performance as a function of increased separation illustrate a deficit in odor pattern separation. Based on these results, it was concluded that lesions of the ventral DG decrease the efficiency of odor based pattern separation, which results in impairments on trials with increased spatial proximity and increased odor similarity among working memory representations (Weeden et al., 2012). The data suggest that the ventral hippocampus, but not dorsal hippocampus, supports pattern separation for odor information.

7. Conclusions

In conclusion, it appears that the subregions (DG andCA1) of the hippocampus play an important role in subserving the process of pattern separation, which is defined as a process to remove redundancy from similar inputs so that events can be separated from each other and interference can be reduced, and in addition can produce a more produce a more orthogonal, sparse, and categorized set of outputs. More specifically it can be shown that the dorsal DG mediates spatial pattern separation, the dorsal CA1 mediates temporal pattern separation for spatial locations and visual objects, the ventral DG mediates odor pattern separation, and the ventral CA1 mediates temporal pattern separation for odors. Similar pattern separation functions can be obtained with animals and humans.

References

Aggleton, J.P., Hunt, P.R., Rawlins, J.N.P., 1986. The effects of hippocampal lesions upon spatial and non-spatial tests of working memory. Behav. Brain Res. 19, 133–146.

- Aimone, J.B., Wiles, J., Gage, F.H., 2006. Potential role for adult neurogenesis in the encoding of time in new memories. Nat. Neurosci. 9, 723–727.
- Amaral, D.G., Witter, M.P., 1995. The hippocampal formation. In: Paxinos, G. (Ed.), The Rat Nervous System. Academic Press, San Diego, pp. 443–493.
- Banks, W.P., 1978. Encoding and processing of symbolic information in comparative judgements. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation: Advances in Theory and Research. Academic Press, New York, pp. 101–159.
- Baumann, O., Chan, E., Mattingley, J.B., 2012. Distinct neural networks underlie encoding of categorical versus coordinate spatial relations during active navigation. Neuroimage 60, 1630–1637.
- Buckmaster, P.S., Schwartzkroin, P.A., 1994. Hippocampal mossy cell function: a speculative view. Hippocampus 4, 393–402.
- Bussey, T.J., Saksida, L.M., Murray, E.A., 2002. Perirhinal cortex resolves feature ambiguity in complex discriminations. Eur. J. Neurosci. 15, 365–374.
- Chiba, A.A., Johnson, D.L., Kesner, R.P., 1992. The effects of lesions of the dorsal hippocampus or the ventral hippocampus on performance of a spatial location order recognition task. Soc. Neurosci. Abstr. 18, 1422.
- Chiba, A.A., Kesner, R.P., Reynolds, A.M., 1994. Memory for spatial location as a function of temporal lag in rats: role of hippocampus and medial prefrontal cortex. Behav. Neural Biol. 61, 123–131.
- Clelland, C.D., et al., 2009. A functional role for adult hippocampal neurogenesis in spatial pattern separation. Science 325, 210–213.
- Cohen,N.J., Eichenbaum H.B., 1993.Memory, amnesia and hippocampal function. Cambridge: MIT press.
- Estes, W.K., 1986. Memory for temporal information. In: Michon, J.A., Jackson, J.L. (Eds.), Time, Mind and Behavior. Springer-Verlag, New York, pp. 151–168.
- Fortin, N.J., Agster, K.L., Eichenbaum, H.B., 2002. Critical role of the hippocampus in memory for sequences of events. Nat. Neurosci. 5, 458–462.
- Gaykema, Ř.P., van der Kuil, J., Hersh, L.B., Luiten, P.G., 1991. Pattern of direct projections from the hippocampus to the medial septum-diagonal band complex: anterograde tracing with Phaseolus vulgaris leucoagglutinin combined with immunohistochemistry of choline acetyltransferase. Neuroscience 43, 349–360.
- Gilbert, P.E., Kesner, R.P., 2002. The amygdala but not the hippocampus is involved in pattern separation based on reward value. Neurobiol. Learn. Mem. 77, 338–353.
- Gilbert, P.E., Kesner, R.P., 2003. Recognition memory for complex visual discrimination is influenced by stimulus interference in rodents with perirhinal cortex damage. Learn. Mem. 10, 525–530.
- Gilbert, P.E., Kesner, R.P., Lee, I., 2001. Dissociating hippocampal subregions: a double dissociation between the dentate gyrus and CA1. Hippocampus 11, 626–636.
- Goodrich-Hunsaker, N.J., Hunsaker, M.R., Kesner, R.P., 2005. Dissociating the role of the parietal cortex and dorsal hippocampus for spatial information processing. Behav. Neurosci. 119, 1307–1315.
- Hannesson, D.K., Howland, J.G., Phillips, A.G., 2004. Interaction between perirhinal and medial prefrontal cortex is required for temporal order but not recognition memory. J. Neurosci. 24, 4596–4603.
- Hasselmo, M.E., Wyble, B.P., 1997. Free recall and recognition in a network model of the hippocampus: simulating effects of scopolamine on human memory function. Behav. Brain Res. 89, 1–34.
- Hasselmo, M.E., Bodelon, C., Wyble, B.P., 2002. A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Comput. 14, 793–817.
- Hoge, J., Kesner, R.P., 2007. Role of CA3 and CA1 subregions of the dorsal hippocampus on the temporal processing of objects. Neurobiol. Learn. Mem. 88, 225–231.
- Holden, H.M., Hoebel, C., Loftis, K., Gilbert, P.E., 2012. Spatial pattern separation in cognitively normal young and older adults. Hippocampus, http://dx.doi.org/ 10.1002/hippo.22017 (advanced online publication).
- Hopkins, R.O., Kesner, R.P., 1993. Memory for Temporal and Spatial Distances for New and Previously Learned Geographical Information in Hypoxic Subjects. Society for Neuroscience, Washington, DC.
- Hopkins, R.O., Kesner, R.P., Goldstein, M., 1995. Item and order recognition memory in subjects with hypoxic brain injury. Brain Cogn. 27, 180–201.
- Hunsaker, M.R., Kesner, R.P., 2008. Evaluating the differential roles of the dorsal dentate gyrus, dorsal CA3, and dorsal CA1 during a temporal ordering for spatial locations task. Hippocampus 17, 955–964.
- Hunsaker, M.R., Rosenberg, J.S., Kesner, R.P., 2008a. The role of the dentate gyrus, CA3a,b, and CA3c for detecting spatial and environmental novelty. Hippocampus 18, 1064–1073.
- Hunsaker, M.R., Fieldsted, P.M., Rosenberg, J.S., Kesner, R.P., 2008b. Dissociating the roles of dorsal and ventral CA1 for the temporal processing of spatial locations, visual objects, and odors. Behav. Neurosci. 122, 643–650.
- Ishizuka, N., Weber, J., Amaral, D.G., 1990. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. J. Comp. Neurol. 295, 580–623.
- Jackson-Smith, P., Kesner, R.P., Chiba, A.A., 1993. Continuous recognition of spatial and nonspatial stimuli in hippocampal lesioned rats. Behav. Neural Biol. 59, 107–119.
- Jeffery, K.J., Anderson, M.I., 2003. Dissociation of the geometric and contextual influences on place cells. Hippocampus 13, 868–872.
- Kesner, R.P., Gilbert, P.E., 2006. The role of the medial caudate nucleus, but not the hippocampus, in a matching-to sample task for a motor response. Eur. J. Neurosci. 23, 1888–1894.
- Kesner, R.P., Bolland, B., Dakis, M., 1993. Memory for spatial locations, motor responses, and objects: triple dissociations among the hippocampus, caudate nucleus and extrastriate visual cortex. Exp. Brain Res. 93, 462–470.

Kesner, R.P., Gilbert, P.E., Barua, L.A., 2002. The role of the hippocampus in memory for the temporal order of a sequence of odors. Behav. Neurosci. 116, 286–290.

Kesner, R.P., Lee, I., Gilbert, P., 2004. A behavioral assessment of hippocampal function based on a subregional analysis. Rev. Neurosci. 15, 333–351.

Kesner, R.P., Hunsaker, M.R., Ziegler, W., 2010. The role of the dorsal CA1 and ventral CA1 in memory for the temporal order of a sequence of odors. Neurobiol. Learn. Mem. 93, 111–116.

- Kirwan, C.B., Stark, E.L., 2007. Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. Learn. Mem. 14, 625–633.
- Leutgeb, J.K., Leutgeb, S., Moser, M.-B., Moser, E.I., 2007. Pattern separation in dentate gyrus and CA3 of the hippocampus. Science 315, 961–966.
- Li, X.-G., Somogyi, P., Ylinen, A., Buzsaki, G., 1994. The hippocampal CA3 network: an in vivo intracellular labeling study. J. Comp. Neurol. 339, 181–208.
- Lisman, J.E., 1999. Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate–CA3 interactions. Neuron 22, 233–242.
- Lorente de Nó, R., 1934. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. J. Psychol. Neurol. 46 (2-3), 113-177.
- Madsen, J., Kesner, R.P., 1995. The temporal-distance effect in subjects with dementia of the Alzheimer type. Alzheimer Dis. Assoc. Disord. 9, 94–100.
- Marr, D., 1971. Simple memory: a theory for archicortex. Proc. R. Soc. B: Biol. Sci. 26, 23–81.
- Marshuetz, C., 2005. Order information in working memory: an integrative review of evidence from brain and behavior. Psychol. Bull. 131, 323–339.
- McDonald, R.J., White, N.M., 1995. Hippocampal and nonhippocampal contributions to place learning in rats. Behav. Neurosci. 109, 579–593.
- Mizumori, S.J., Ragozzino, K.E., Cooper, B.G., Leutgeb, S., 1999. Hippocampal representational organization and spatial context. Hippocampus 9, 444–451.
- Morris, A.M., Churchwell, J.C., Kesner, R.P., Gilbert, P.E., 2011. Selective lesions of the dentate gyrus produce disruptions in place learning for adjacent spatial locations. Neurobiol. Learn. Mem. 97, 326–331.
- Mumby, D.G., Wood, E.R., Pinel, J.P.J., 1992. Object recognition memory is only mildly impaired in rats with lesions of the hippocampus and amygdala. Psychobiology 20, 18–27.
- Myers, C.E., Gluck, M.A., Granger, R., 1995. Dissociation of hippocampal and entorhinal function in associative learning: a computational approach. Psychobiology 23, 116–138.
- Norman, G., Eacott, M.J., 2004. Impaired recognition with increasing levels of feature ambiguity in rats with perirhinal cortex lesions. Behav. Brain Res. 148, 79–91.
- O'Keefe, J., Burgess, N., 1996. Geometric determinants of the place field of hippocampal neurons. Nature 381, 425–428.
- O'Reilly, R.C., McClelland, J.L., 1994. Hippocampal conjunctive encoding, storage, and recall: avoiding a trade-off. Hippocampus 4, 661–682.
- Otto, T., Eichenbaum, H., 1992. Complementary roles of the orbital prefrontal cortex and the perirhinal–entorhinal cortices in an odor-guided delayed-nonmatchingto-sample task. Behav. Neurosci. 106, 762–775.

- Poucet, B., 1989. Object exploration, habituation, and response to a spatial change in rats following septal or medial frontal cortical damage. Behav. Neurosci. 103, 1009–1016.
- Risold, P.Y., Swanson, L.W., 1997. Connections of the rat lateral septum. Brain Res. Rev. 24, 115–195.
- Rolls, E.T., 1989. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Bryne, J.H., Berry, W.O. (Eds.), Neural Models of Plasticity: Experimental and Theoretical Approaches. Academic Press, San Diego, pp. 240–265.
- Rolls, E.T., 1996. A theory of hippocampal function in memory. Hippocampus 6, 601–620.
- Rolls, E.T., Kesner, R.P., 2006. A computational theory of hippocampal function, and empirical tests of the theory. Prog. Neurobiol. 79, 1–48.
- Saksida, L.M., Bussey, T.J., Buckmaster, C.A., Murray, E.A., 2006. No effect of hippocampal lesions on perirhinal cortex-dependent feature-ambiguous visual discriminations. Hippocampus 16, 421–430.
- Samsonovich, A., McNaughton, B.L., 1997. Path integration and cognitive mapping in a continuous attractor neural network model. J. Neurosci. 17, 5900–5920.
- Scharfman, H.E., 2007. The CA3 "backprojection" to the dentate gyrus. In: Scharfman, H.E. (Ed.), The Dentate Gyrus: A Comprehensive Guide to Structure, Functional and Clinical Implications. Progress in Brain Research, vol. 163, pp. 627–638.
- Shapiro, M.L., Olton, D.S., 1994. Hippocampal function and interference. In: Schacter, D.L., Tulving, E. (Eds.), Memory Systems 1994. MIT Press, London, pp. 141–146.
- Small, S.A., Schobel, S.A., Buxton, R.B., Witter, M.P., Barnes, C.A., 2011. A pathophysiological framework of hippocampal dysfunction in ageing and disease. Nat. Rev. Neurosci. 12, 585–601.
- Stark, S.M., Yassa, M.A., Stark, C.E., 2010. Individual differences in spatial pattern separation performance associated with healthy aging humans. Learn. Mem. 17, 284–288.
- Tanila, H., 1999. Hippocampal place cells can develop distinct representations of two visually identical environments. Hippocampus 9, 235–246.
- Tolentino, J.C., Pirogovsky, E., Luu, T., Toner, C.K., Gilbert, P.E., 2012. The effect of interference on temporal order memory for random and fixed sequences in nondemented older adults. Learn. Mem. 19, 1–6.
- Treves, A., Rolls, E.T., 1994. Computational analysis of the role of the hippocampus in memory. Hippocampus 4, 374–391.
- Weeden, C.S.S., Hu, N.J., Ho, L.Y.N., Kesner, R.P., 2012. The role of the ventral dentate gyrus in olfactory learning and memory. Soc. Neurosci. Abstr.
- Witter, M.P., 1993. Organization of the entorhinal-hippocampal system: a review of current anatomical data. Hippocampus 3, 33–44.
- Wright, A.A., 2007. An experimental analysis of memory processing. J. Exp. Anal. Behav. 88, 404–433.
- Yassa, M.A., Stark, C.E.L., 2011. Pattern separation in the hippocampus. Trends Neurosci. 34, 515–525.
- Yassa, M.A., Lacy, J.W., Stark, S.M., Albert, M.S., Gallagher, M., Stark, C.E.L., 2010. Pattern deficits associated with increased hippocampal CA3 and dentate gyrus in nondemented older adults. Hippocampus 21, 968–979.

Contents lists available at SciVerse ScienceDirect





Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Jocelyne Bachevalier*, Anthony A. Wright, Jeffrey S. Katz¹

Department of Anatomy and Neurobiology, University of Texas, Health Science Center-Houston, TX, United States

ARTICLE INFO

Article history: Received 2 August 2012 Received in revised form 12 November 2012 Accepted 26 November 2012

Keywords: Recognition memory Primacy Recency Serial list memory Proactive interference

ABSTRACT

We examined the role of the hippocampus in list-memory processing. Three rhesus monkeys that had extensive experience in this task and had demonstrated full abstract-concept learning and excellent list memory performance (Katz et al., 2002; Wright et al., 2003) received bilateral neurotoxic hippocampal lesions and were re-tested in the serial list memory task. Effects of delays on memory performance were measured in all monkeys, whereas the effects of proactive interference were assessed in only one. Despite a slight change in performance of one of the three animals during re-learning of the same/different task, selective hippocampal damage had little or no effects on list memory accuracy. In addition, the hippocampal damage did not impact serial list position functions (SPFs) but slightly altered the dynamic of the SPF curves. Finally, even more remarkable was that accurate memory performance. Together the findings indicate that, with short list items and extensive training in the task (i.e., reference memory), monkeys with selective hippocampal lesions may be able to use alternative memory processes (i.e., working memory) that are mediated by structures other than the hippocampus.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decade, the contributions of the hippocampus and medial temporal lobe cortex to recognition have generated a host of studies in many species, but at the current time the specific role of each of these brain structures remains heavily debated. An example is provided by the recent issue of the journal "Hippocampus" (2010, vol. 20) exposing the different views fueling this debate. One of the reasons this controversy has lasted so long is the disagreement over how to measure hippocampal and cortical contributions to recognition memory. Recognition memory in humans is commonly assessed with list learning tasks in which participants study a set of stimuli (pictures of objects, visual patterns, faces, or words), and after a delay, judge whether the stimuli are familiar (included in the list) or new. Studies on human amnesic patients with damage to the hippocampus or adjacent cortical areas (Aggleton and Shaw, 1996; Bowles et al., 2007; Holdstock et al., 2002; Mayes et al., 2003; Mishkin et al., 1998; Reed et al., 1997; Stark et al., 2002; Vargha-Khadem et al., 1997) and functional imaging studies (Yonelinas and Parks, 2007; for reviews, see Eichenbaum et al., 2007; Skinner and Fernandes, 2007; Wais, 2008) have suggested that the hippocampus is involved in recognition memory only when participants fully recollect the items (i.e., the items and all other information associated with the items, such as whether the words were shown in red or green or the pictures were emotionally positive or negative), but not when they simply used familiarity judgments (was the item in the list or not?), which are supported by the medial temporal cortical areas. Another view, however, proposed that the strength of the memory traces is the critical attribute such that memory traces with strong or weak memory load may require the hippocampus and medial temporal cortex, respectively (Squire et al., 2007; Wixted et al., 2010).

Animal studies have attempted to resolve this disagreement but without convincing success so far. For example in monkeys, recognition memory has generally been investigated using delayed matching-to-sample (DMTS) or delayed nonmatching-to-sample (DNMTS) tasks in which the animal has to indicate which of two stimuli has been seen earlier by choosing either the familiar (match) or the novel (nonmatch) stimuli presented together

 $[\]stackrel{ au}{=}$ Jonathon Crystal was the acting editor for this paper.

^{*} Corresponding author at: Yerkes National Primate Research Center and Psychology Department, Emory University, 954 Gatewood Road, Atlanta, GA 30329, United States. Tel.: +1 404 727 9765; fax: +1 404 727 8088.

E-mail address: jbachev@emory.edu (J. Bachevalier).

¹ Now at Department of Psychology, Auburn University, 226 Thach Hall, Auburn, AL 36849, United States.

^{0376-6357/\$ -} see front matter © 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.beproc.2012.11.012

during a choice test. Generally, these tasks employed a large pool of stimuli (500 to thousands). Memory is then further assessed by increasing the delays between the sample presentation and the choice or by increasing the list of items to be remembered. Using these tasks, lesion studies have provided conflicting results. Thus, whereas some studies have reported recognition deficits at the long delays or long lists following selective hippocampal lesions (Beason-Held et al., 1999; Zola et al., 2000), others found no impairment (Baxter and Murray, 2001; Murray and Mishkin, 1998; Nemanic et al., 2004). One potential limitation with the nonhuman primate studies is that the DMTS and DNMTS paradigms may rely on memory processes different from those that support the list memory tasks in humans (see Nemanic et al., 2004). The memory processes supporting DNMTS performance could include familiarity judgment, working memory, or retrospective processing, which could recruit brain areas other than the hippocampus, such as the medial temporal and prefrontal cortices known to be critical for normal performance on DNMTS tasks (Bachevalier and Mishkin, 1986; Brown and Aggleton, 2001; Ennaceur et al., 1996; Fahy et al., 1993; Gaffan and Murray, 1992; Kolb et al., 1994; Meunier et al., 1993; Miller et al., 1996; Murray and Bussey, 1999; Nemanic et al., 2004; Pihlajamaki et al., 2004; Simons and Spiers, 2003; Suzuki et al., 1993; Xiang and Brown, 2004).

In an attempt to investigate further the reasons for this disagreement and enable better comparisons with results from the human literature, the present study employed a serial list memory task similar to that used in humans (Wright et al., 1985) to re-assess the effects of selective hippocampal lesions on recognition memory in monkeys. In this task, animals are presented with a short list of items on a computer monitor followed by a probe test. The probe test presents either an item seen in the list or a new item together with a white rectangle. To receive a reward, the animal has to touch the item on the screen if it was an item of the list or touch the white rectangle if the item was new. The serial list memory task offers several advantages relative to the previous matching tasks. First, as compared to the DNMTS task in which both the familiar and new items are present together on the screen during the animal's selection (familiar versus novel), the serial list memory task presents only one item necessitating a "yes/no" or "same/different" response. Thus, the forced-choice response in the DNMTS task may favor the use of familiarity/novelty judgment that are more dependent upon the medial temporal cortex to the detriment of same/different relational representations and retrieval strategies, which depend more heavily upon the hippocampus (Damasio et al., 1985; Eichenbaum et al., 1989, 2007; O'Keefe and Nadel, 1978; Rudy and Sutherland, 1989, 1992; Shapiro and Olton, 1994; Sutherland and Rudy, 1989).

Another important advantage of the serial list memory task over the DNMTS in the investigation of the participation of the hippocampus in recognition memory is that the list memory task can better dissociate different memory processes. In a previous study comparing serial list memory abilities in pigeons, monkeys and humans, Wright et al. (1985) demonstrated that the typical serial U-shaped position function with good (long-term) memory of the first list items (primacy effect) and a good (short-term) memory of the last list items (recency effect) normally found in human studies was also present for pigeons and monkeys. Furthermore, the authors demonstrated that in those three species, the shape of the serial-position function changed with varying the retention intervals between the end of the list and the probe test. That is, at short retention delays, recognition memory increased monotonically with better memory for the last items of the list; for intermediate delays, the serial list curve had U-shape functions with better memory for the first and last items than for the middle ones; lastly, for long retention delays, recognition memory decreased monotonically with better memory for the first items of the list. The authors suggested that these dynamic changes in serial-position functions reflect the participation of two or more memory processes. This conclusion is strengthened by the numerous demonstrations showing that the primacy and recency effects can be independently altered. Variables that selectively affect the recency effect include: moderate to long retention delays (e.g., Gardiner, 1974; Glanzer and Cunitz, 1966; Postman and Phillips, 1965; Roediger and Crowder, 1976; Wright et al., 1985); auditory vs. visual modality of stimulus presentation (e.g., Crowder, 1986; Crowder and Morton, 1969; Murdock, 1966; Wright, 2007); and knowledge about the end of the list (Watkins and Watkins, 1974). Variables that selectively affect the primacy effect include: fast presentation rates (Glanzer and Cunitz, 1966), long list lengths (Murdock, 1962), very short retention delays in single-item recognition tasks (Wright et al., 1985), alcohol intoxication (Jones, 1973), and mental retardation (Belmont and Butterfield, 1971).

Interestingly, there exists also neuropsychological evidence to support this functional dissociation of memory processes in serial list learning task. Thus, different brain areas seem to independently support the primacy and recency effects. The prefrontal cortex known to be critical for working memory processes and perirhinal cortex known to mediate short term memory have been associated with the recency effect (Barker and Warburton, 2011; Goldman-Rakic, 1987; Kesner, 1985; Saffran and Marin, 1975; Warrington et al., 1971; Warrington and Shallice, 1984; Weiskrantz, 1987), whereas the hippocampus has been associated with long-term (primacy) memory (e.g., Baddeley and Warrington, 1970; Hermann et al., 1996; Hopkins and Kesner, 1995; Hopkins et al., 1995; Kesner, 1998; Kesner and Novak, 1982).

The advantages provided by the serial list learning task over the DMTS and DNMTS offers an improved method with which to assess the role of hippocampus in recognition memory. More importantly, task manipulations, such as length of the delays and magnitude of the proactive interference across list items, may inform recent theories concerning the precise role of the hippocampus in recognition memory (see reviews of the current neural models in the review Hippocampus, 2010, vol. 20). Therefore, in this study, three rhesus monkeys with extensive experience in a serial list memory task were used (Katz et al., 2002; Wright et al., 2003). All monkeys had demonstrated full abstract-concept learning and excellent list memory performance before receiving bilateral neurotoxic lesions of the hippocampal formation. After recovery from surgical procedures the monkeys were then re-tested in the serial list learning task.

2. General methods

All procedures were approved by the Animal Care and Use Committee of the University of Texas Health Science Center at Houston in Houston, TX and carried out in accordance with the National Institute of Health Guide for the Care and Use of Laboratory Animals. All efforts were made to minimize the number of animals used, as well as any pain and suffering.

2.1. Subjects

Subjects were three, 6–12 year-old, rhesus monkeys (*Macaca mulatta*) of both sexes, weighing 5–12 kg (Cuba, Gracie, and Slim). They were housed individually and maintained on a 12:12 h light–dark cycle. Multi-vitamins were given daily and fresh fruit weekly. All three monkeys received presurgical training on a twoitem same/different task and then list memory (Katz et al., 2002; Wright et al., 2003). Experimental training sessions were conducted 5–7 days a week. On testing days, access to food (Purina Monkey Chow) and water in their home cages was restricted about 15 h before testing. On non-testing days, they had accessed to food and water ad libitum. Additionally, vegetable and fruit supplements were provided at the end of each week. All monkeys were first trained on the same/different task and then received a transfer test to assess same/different concept learning, followed by a list memory task to measure serial position functions. After this training, all animals received bilateral neurotoxic (ibotenic acid) lesions of the hippocampal formation. After recovery from the surgical procedure, they were re-trained on the list memory task, as described below.

2.2. Apparatus

2.2.1. Chamber

Monkeys were tested unrestrained in one of two identical $(47.5 \text{ cm wide} \times 53.13 \text{ cm deep} \times 66.25 \text{ cm high})$ custom aluminum test chambers (see Bhatt and Wright, 1992 for further details). A fan (Dayton 4C440, Niles, IL) located in the ceiling of the chamber, provided ventilation and white noise. Bio-Serv banana pellets (300 mg) were dispensed (model # ENV-203-300 MED Associates, Inc., Georgia, VT) into a pellet cup (5.6 cm diameter and 2.5 cm deep) that was 10 cm from the left edge and 52.5 cm from the top edge of the front panel. Tang orange drink was delivered (model # 71215 Honeywell, Inc., New Britain, CT) through a juice spout that was 10 cm from the right edge and 42.5 cm from the top edge of the front panel. Touches to a computer monitor were detected by an infrared touch screen bezel (model # 81009703-01 Carroll Touch, Round Rock, TX). The bezel fitted snugly within a $40 \text{ cm} \times 33.75 \text{ cm}$ cut-out in the front panel that was centered 9.38 cm from the top of the operant chamber. Touch responses were directed by a Plexiglas template $(32.5 \text{ cm} \times 40 \text{ cm})$ with cut-outs matching the size and location of the stimuli.

2.2.2. Stimuli

Travel-slide color pictures were digitized via a Howtek Photomaster (87RU) camera and a Truevision TARGA-16 processing card in a 256 × 256 resolution. Stimuli were presented on a 39 cm color monitor (Eizo FX-C6; Ishikawa, Japan, 600 × 480-pixel resolution). Stimulus displays consisted of two travel-slide color pictures (each 13.75 cm × 9.7 cm) and a white rectangle (6.25 cm × 5.6 cm) on a black background. The pictures were vertically aligned with a 3.4 cm gap between them. The top picture was centered 20.63 cm from the left edge and 18.75 cm from the top of the front panel. The bottom of the white rectangle was horizontally aligned with the bottom of the lower picture with a 3.7 cm gap between them. A set of 432 travel slides was used in these experiments (for the complete set of stimuli see Wright and Katz, 2006).

2.2.3. Experimental control

Experimental events were controlled and recorded via custom software written in Visual Basic on a Pentium personal computer. A video card (Appian Jeronimo J2 Advanced Graphics Accelerator) controlled the monitor. A computer-controlled relay interface (model # PI0-12, Metrabyte, Taunton, MA) operated the juice valve and pellet dispenser.

2.3. Pre-surgical training procedures

Detailed description of data obtained on presurgical training procedures has been published (Katz et al., 2002; Wright et al., 2003). Table 1 summarizes the presurgical training history for each monkey as well as the number of sessions they received for each phase.

Briefly, monkeys were first trained to retrieve banana pellets (Noyes 300 mg) from the food cup and drink Tang orange juice from the juice spout. Responses to the video monitor and touch

Table 1

Number of sessions received for each training phase prior to surgery. Session consisted of 100 trials and Cuba started training with 10 touches to the sample.

Monkey	Gracie	Slim	Cuba
Training			
0-Response	250	250	NA
Fixed-time	54	30	NA
10-Responses	0	26	35
Transfer training	41	46	33
List-length testing	31	37	45
Delays testing	205	995	497

screen were shaped by successive approximations to a white rectangle positioned at the right bottom corner of the screen (later to become the different response area) and another white rectangle (13.75 cm \times 9.7 cm) placed in the position where the lower picture will appear on each trial. These two rectangles appeared on separate trials and were randomly and equally presented over a 100-trial session. A single touch to the white rectangle was followed by a 1.1 s, 660 Hz tone. The rewards were either food (1 pellet, delivered simultaneously with the tone) or juice (3–5 cm³, delivered 1 s after the tone) and a 15 s intertrial interval (ITI) followed reinforcement. Once a monkey was consistently responding (1–5 sessions), same/different (S/D) training began.

2.3.1. Same/different training

Eight pictures were used for S/D training. They were arranged in 64 pairs (8 "same" and 56 "different") and each session consisted of 100 trials (50 "same"/50 "different"). For Gracie and Slim, a trial began with presentation of the two pictures (one above the other) and the white rectangle. If the two pictures were the same, a touch response to the lower picture was rewarded. If the two pictures were different, a touch response to the white rectangle was rewarded. For Cuba, a trial began with presentation of the upper picture (sample stimulus) only. Initially, the monkey was required to touch the picture once, before the two pictures appeared on the screen. Then, the number of touches to the sample picture was progressively increased to a maximum of 10 touches. Only correct choices were rewarded and starting on the fifth training session, incorrect choices were followed by a repeat of the incorrect trial (correction procedure). Training continued until performance was 80% or better on three consecutive sessions. The correction procedure was then removed and training continued until the same criterion was met. Whereas Cuba learned the task in 35 sessions (100 trials/session), neither Gracie nor Slim learned the same/different task in the limit of testing, i.e., 250 sessions (Katz et al., 2002). Slight modifications were done on the task to allow Gracie and Slim to learn the same/different task. Both monkeys received additional sessions (54 and 30 sessions, respectively) in which the sample stimulus remained on the screen for a fixed-time period before the choice test was presented. The length of time that the upper picture remained on the screen was made equivalent to the average time required by Cuba and two other monkeys (see Katz et al., 2002) to touch 10 times the sample picture. Gracie learned the S/D with the fixed-time procedure, but Slim did not. Slim received an additional 26 sessions during which he was required to touch 10 times the sample stimulus before the choice test was presented (as did Cuba).

2.3.2. Transfer testing

Following S/D training, all three monkeys received six consecutive transfer test sessions to assess abstract-concept learning. Like training, each transfer session had 100 trials. They received 10 transfer trials (5 "same" and 5 "different") consisting of novel stimuli pseudo-randomly intermixed with 90 baseline training trials (45 "same" and 45 "different") consisting of the same 8 pictures used in the S/D training. The set size was then progressively increased from 8 to 16, 32, 64, and 128 pictures. Transfer tests were conducted after reaching the performance criterion following the 32, 64, and 128 set sizes.

2.3.3. List-length and retention delay training

List-length training began with one list item. The upper image was presented and after 10 touches, the list item disappeared for 1 s, following which the lower picture and the white rectangle appeared on the screen. Monkeys received 100 trials per session until reaching the criterion of 85% correct or better. The list was then extended progressively to 2, 3, 4, 6, 8 and 10 items. The interstimulus interval (ISI) and the delay between the last item of the list and the probe item were 1 s. The training set size was increased from 128 items to as many as 432 items during this training.

Following list-length expansion training, monkeys were returned to training and testing with a list length of 4 items. Again list items were presented for 1 s, 1 s ISI, and a 1 s retention delay was interposed between the last list item and the probe item. After performance stabilized with this procedure, retention delays of 0, 1, 2, 10, 20 and 30 s were tested in a block design. Each block consisted of six sessions of 32 trials (16 same/16 different) each with the retention delay fixed. The serial position of the matching picture on same trials was counterbalanced and randomized within each session. Two sessions, with different retention delays, were tested daily. One delay was short (0, 1 and 2 s) and one was long (10, 20 and 30 s). The order (short, long) was counterbalanced over successive six session blocks. Each delay was selected randomly with the constraint that each delay occurred once within a block. Training items were unique in a session. Training continued until performance was accurate and the serial position functions were stable over several days of testing.

3. Experiment 1

To investigate the role of the hippocampal formation on the retention of S/D abstract-concept learning and on the shape and time-course of serial position functions, the three animals received MRI-guided neurotoxic lesions of the hippocampal formation bilaterally, and, following recovery, were re-tested on the "S/D" and list memory tasks.

3.1. Methods

3.1.1. Neuroimaging procedures

Briefly, one to three weeks prior to surgery, each subject was placed in a non-ferromagnetic stereotaxic device (Crist Instruments, Co., Inc., Damascus, MD), and received a T1-weighted, structural MRI scan through the entire brain at 1 mm slice intervals in the coronal plane (see Nemanic et al., 2002 for details). These high-resolution images were used to create an individual atlas for each monkey from which the coordinate values for each neurotoxin injection site for the hippocampal lesion were calculated.

The hippocampal lesions were produced by injections of ibotenic acid and were intended to include all ammonic fields, the dentate gyrus, and subicular complex. For each monkey, MRI coordinates (i.e., anterior-posterior, medial-lateral, dorsal-ventral) of the injection targets were taken through the entire rostral-caudal extent of the structure. For the posterior two-thirds of the hippocampal formation, one coordinate was selected per MR image (every millimeter) to target the center of the hippocampus body. For the most anterior portion of the hippocampus, where the uncus was clearly visible, two coordinates were taken per image, one to target the body of the hippocampal formation (lateral) and one to target the uncus (medial). The MRI coordinates for the selected injection sites were then converted into stereotaxic coordinates relative to the stereotaxic point zero. A total of 11–12 injection sites were selected per hippocampus, 9–10 sites were spaced 1.5 mm apart, through the body of the hippocampal formation, and 2 additional sites were spaced 1.5 mm apart, at the uncus.

3.1.2. Surgery

All surgical procedures were carried out under deep anesthesia (7:3 mixture of ketamine hydrochloride, 100 mg/ml, and robinul 0.2 mg/ml, 0.1 ml/kg i.m.), followed by isoflurane inhalation (2% to effect) using aseptic conditions. The anesthetized animal was repositioned in the stereotaxic apparatus for the hippocampal lesions. Heart and respiration rates, blood pressure, expired CO₂, and body temperature were monitored throughout the procedure.

Following disinfection (Nolvasan solution) of the scalp and application of local anesthetic (Marcaine 25%, 1.5 ml) along the incision line, the skin was cut from the orbit to the occiput, and the connective tissue and temporal muscles were gently retracted and the dura cut.

To access the injection sites, a bone flap was made on the top of the skull and small slits were cut in the dura over the location of the injection sites. Injections were performed simultaneously in the left and right hemispheres. The needle of the 10 μ l Hamilton syringe, held in a Kopf electrode manipulator (David Kopf Instruments, Tujunga, CA), was slowly lowered to each injection site and a total of 1.5–2.4 μ l of ibotenic acid (Regis Chemical, Morton Grove, IL, 15 mg/ml in PBS, pH 7.4) was injected at each site at a rate of 0.2 μ l/min. A 5 min delay was imposed before retraction of the needle to permit diffusion of the neurotoxin and minimize its spread along the needle track. The needle was then swabbed to remove any residual neurotoxin, and repositioned and lowered to the next injection coordinate.

When the injections were completed, the dura openings were sewn, and all tissues were closed in anatomical layers. To minimize brain swelling, all operated animals received an intravenous drip of 30 ml of mannitol (20%, delivered at a rate of 1 ml/min) before beginning the final injection.

Beginning 12 h prior to and continuing for one week after surgery, all operated monkeys were treated with dexamethazone sodium phosphate (0.4 mg/kg, i.m.) and Cephazolin (Bristol-Myers Squib, 25 mg/kg, i.m.) to reduce inflammation and protect against infection, respectively. For 3 days after surgery, the monkeys also received an analgesic (acetaminophen 10 mg/kg, p.o.).

3.1.3. Lesion assessment

Seven to ten days after surgery, all monkeys received a second MRI procedure, including a high resolution T1 and a Fluid Attenuated Inversion Recovery (FLAIR) scans to visualize areas of increased water density induced by cell death at the injection sites. The post-surgical scanning procedures have been shown to provide a rapid and accurate way to quantify lesion location and extent after neurotoxin infusion (Málková et al., 2001; Nemanic et al., 2002). In one case (Gracie), the extent of damage was minimal on the left side and so a second surgery was performed 1 month later during which ibotenic acid was injected unilaterally at 8 sites along the spared hippocampus. A second FLAIR, one week later, confirmed the additional damage by the ibotenic acid to the left hippocampus.

All three monkeys died during Tropical Storm Allison, thus, hypersignals observed onto post-surgical FLAIR images were the only means to examine the location and extent of damaged areas. However, estimation of the extent of hypersignals was shown to provide a good estimate of the extent of cell loss observed histologically and to correlate with the amount of volume reduction observed months after surgery (r=0.893, p<0.005; Málková et al., 2001; Nemanic et al., 2002). For each animal, FLAIR images were matched with digitized drawings of coronal sections at

1 mm intervals through a normal monkey brain. The extent of hypersignals on each image was drawn on each coronal section of the normal monkey. Using Scion Image[®] software, the extent of hypersignals seen within the hippocampal formation as well as in adjacent structures was measured on each section and percent damage for each structure (as compared to the normal brain) was calculated. For Gracie, which received ibotenic acid injections in two stages in the left hemisphere, the extent of hypersignals seen on the 2 post-surgical scans were combined and drawn onto the coronal sections of the normal brain.

3.1.4. Postsurgical training

Following recovery from surgery, which lasted 40, 49 and 52 days for Cuba, Gracie and Slim, respectively, all monkeys were re-trained on the list learning task, using the same procedures described above for the pre-surgical training. The animals were first given training S/D sessions followed by three transfer sessions to assess retention of abstract-concept learning. Training sessions consisted of 32 trials (16 same, 16 different) with stimuli randomly selected for each trial from a set size of 432 images. Upon completing an FR 1 to the upper image, it disappeared, followed by a 1 s retention delay and the lower image. Transfer sessions were identical to training sessions except 16 transfer trials (8 same, 8 different) replaced baseline trial. Slim was not given the transfer sessions at this time because of poor performance in the S/D task, but he was tested later, after completing list memory testing. Next, two monkeys (Cuba and Gracie) were returned to list memory testing using a list of 4 items in the block design previously described. Testing continued in this design until the serial position functions were stable. Slim also returned to list memory testing but with short delays (0, 1 and 2 s) because of his difficulty in the S/D task. After 24 sessions (8 of each short delay) his performance improved and he began training in the block design with all six retention delays.

3.2. Results

3.2.1. Lesion extent

Examples of post-surgical FLAIR images matched with corresponding coronal sections of the normal monkey are shown for the three cases in Figs. 1–3. The percent damage to each field of the hippocampal formation for each hemisphere in each case is presented in Table 2.

Damage to the hippocampal formation was incomplete, but substantial, in all cases, ranging from 57% to 97%. In one case, Cuba, cell loss was extensive, encompassing all but the rostral most 2 mm of the hippocampal formations on the right (see Fig. 1, level +12), whereas in the others, Gracie (Fig. 2) and Slim (Fig. 3), cell loss was extensive (>85%) on one side, but only moderate on the other (<43%). Nevertheless, despite an incomplete lesion on the left hemisphere, Gracie had 71% and 91% damage to the CA1 and CA2 fields, respectively. Importantly, none of the three monkeys had unintended damage to the cortical areas surrounding the hippocampal formation. Slim's lesions, however, encroached on the posterior amygdala bilaterally (12% and 9% on the right and left, respectively). In addition, Slim incurred damage to the head of the caudate and tail of the putamen (more left than right hemisphere), resulting most likely from the injection needles.

3.2.2. Re-training and transfer on the same/different task

Two monkeys (Gracie and Cuba) showed excellent retention of the S/D task they had learned pre-surgically and were retrained for only two sessions (64 trials) averaging 93.75% and 81.3%, respectively. By contrast, Slim performed more poorly averaging only 55% over 11 sessions (352 trials). Therefore, Gracie and Cuba immediately proceeded to transfer testing and showed full abstract-concept learning again (Gracie: baseline = 98%, CUBA



Fig. 1. FLAIR MR coronal images (top) matched with drawing sections of a normal monkey brain (bottom) for three A–P levels through the hippocampal formation in CUBA. White areas on the FLAIR images depict areas of hypersignals that were reconstructed onto the corresponding section (gray area) of the normal monkey brain. MR acquisition parameters for the three scans were: slice thickness: 3.0 mm (with 1 mm offset between scans), repetition time: 10, echo time: 1.47, inversion time: 2.2, number of averages: 2.0, echo numbers(s): 1, magnetic field strength: 1.5 T, spacing between slices: 3 mm echo train length: 24. The numeral on the left indicates the distance in millimeters from the interaural plane. *Abbreviations*: CA1, CA2, and CA3, cornu ammonis fields of the hippocampus; DG, dentate gyrus; ERh, entorhinal area 28; HATA, amygdala–hippocampus; TE, TEO, TH, and TF: cytoarchitectonic fields described by von Bonin and Bailey (1947), and V2, visual extrastriate cortical area.

transfer = 98%; Cuba: baseline = 98%, transfer = 95.8%). For Slim, after completing list memory testing, he also showed full abstract-concept learning again (baseline = 89.6% and transfer = 91.7%).

3.2.3. List memory

Fig. 4 illustrates serial position functions (SPFs) obtained during the last 5 blocks of pre-surgical and post-surgical list-memory testing for each monkey. At the end of post-surgical list-memory training, all animals showed high accuracy on the list memory task with some changes in the shape of the SPFs compared to presurgical SPFs. For Gracie (Fig. 4, top), the primacy effect during the pre-surgery training only emerged after a 10 s delay, whereas the primacy effect during post-surgery was present at all delays,

Table 2				
Percent of i	ntended damage	e to the hip	pocampal	formation. ^a

Subjects	CA1			CA2			CA3					
	L	R	Avg	W	L	R	Avg	W	L	R	Avg	W
Cuba	80.9	100	90.5	80.9	61.7	100	80.9	61.7	58.9	97.9	78.4	57.7
Gracie	71.2	100	85.6	71.2	91.4	92.2	91.8	84.2	34.6	89.9	62.2	31.1
Slim	94.1	5.7	49.9	5.4	90.1	7.7	48.9	6.9	95.2	32.3	63.7	30.7
Χ	82.1	68.6	75.3	52.5	81.1	66.6	73.9	50.9	62.9	73.4	68.1	39.8
Subjects	Dentate	gyrus			Subicula	r complex			Total			
	L	R	Avg	W	L	R	Avg	W	L	R	Avg	W
Cuba	66.6	96.6	81.6	64.4	66.5	94.0	80.2	62.5	66.2	97.4	81.8	64.5
Gracie	34.0	84.3	59.2	28.7	16.9	84.4	50.6	14.2	43.3	85.7	64.5	37.1
Slim	78.6	35.7	57.1	28.0	89.8	23.5	56.7	21.1	88.2	25.9	57.1	22.9
Х	59.7	72.2	65.9	40.4	57.7	67.3	62.5	32.6	65.9	69.7	67.8	41.5

^a Data are estimated intended damage (in percent of normal) to the ammon fields (CA1, CA2, and CA3), the dentate gyrus and subicular complex. Total refers to average of all five hippocampal regions. Note that for Gracie, percent of damage included extent of hypersignals found after both the first and second surgical procedures. *Abbreviations*: L, percent damage to the left hemisphere; R, percent damage to the right hemisphere; Avg, average of *L* and *R*, $W = (L \times R)/100$ (weighted index as defined by Hodos and Bobko, 1984).



SLIM



Fig. 2. FLAIR MR coronal images (top) matched with drawing sections of a normal monkey brain (bottom) for three A–P levels through the hippocampal formation in GRACIE. Note that the MR images shown were those obtained after the first surgical procedure; note the extensive sparing of the hippocampal formation on the left hemisphere as reproduced on the drawing below. Conventions as in Fig. 1.

Fig. 3. FLAIR MR coronal images (top) matched with drawing sections of a normal monkey brain (bottom) for three A–P levels through the hippocampal formation in SLIM. Conventions as in Fig. 1.

including those shorter than 10 s. In addition, the recency effect dissipated at a slightly shorter delay (2 s) after surgery as compared to pre-surgery (10 s). Slim, like Gracie, also showed an earlier primacy effect (0 s delay) post-surgery than pre-surgery (Fig. 4, center), but the primacy effect grew more slowly with increasing delays than for Gracie. Slim did not show dissipation of the recency effect (presurgery or post-surgery) at longer delays, but his good accuracy was somewhat confounded by a bias to respond "same" at long delays. Monkey Cuba was clearly the most exceptional performer both preand post-surgically (Fig. 4, bottom). Cuba's primacy effect remained very strong across all delays and interestingly Cuba's post-surgery recency effect did not dissipate at long delays (even at the longest delay of 30 s) as it had pre-surgically.

3.2.4. Extended delay training

Given the unexpected failure to identify important changes in memory performance at delays up to 30 s with compromised or absent hippocampi, we extended the delays of Cuba and Slim further as shown in Fig. 5. Cuba was tested with longer delays of 60, 120, and 240 s and Slim with longer delays of 60 and 120 s (along with the 0, 1 and 2 s short delays) in the block design previously described. Short-delay performance was basically unaltered for these two monkeys, and remarkable was that their longer delay performance was basically unchanged from what it had been with the 30 s delay. Cuba's performance averaged 84%, 82%, and 93% for long delays of 60, 120, and 240 s, and Slim's performance averaged 78% and 89% for long delays of 60 and 120 s. In general, the longer delays had very little effect on their overall list memory performance and only slightly affected the form of some of their serial position functions.

4. Experiment 2

The accurate performance and regular serial position functions from Experiment 1, even when the delays between the end of the list and the probe test were increased, suggest that this good list-memory performance was not dependent upon the hippocampus. The majority of the serial position functions for each animal remained similar to those they obtained prior to the hippocampal lesions, with the only changes being improved primacy performance for Gracie and Slim at short delays, and improved recency performance for Cuba at long delays.

Indeed, the complete lack of decrements in memory performance following hippocampal lesions on a S/D recognition task in fact parallels many of the findings reported in humans with temporal lobe amnesia using yes/no recognition tasks (Bastin et al., 2004; Freed et al., 1987; Holdstock et al., 2005; Mayes et al., 2002). One reason for this lack of effects might be that these tasks may be solved using familiarity judgments "Have I seen this item before or not?" mediated by the medial temporal cortex rather than by accurate recollection about "when" or "where" (contexts) that this item ("what") occurred. Such context learning is the basis for episodic memory and has been concluded to depend on the hippocampus (see for review Brown and Aggleton, 2001; Eichenbaum et al., 2007; Norman and O'Reilly, 2003; Rugg and Yonelinas, 2003). According to these theorists, in the absence of a functional hippocampus, serial list memory of the three monkeys could only have been maintained at a high level of proficiency by the use of familiarity judgments possibly mediated by the medial temporal cortical areas.

One way to test this possibility of familiarity processing would be to reduce the stimulus set size from the original large pool of items (used only once in a daily session) to a set size of 8 items that are repeated many times in different lists within the daily session. Thus, simple familiarity "Have I seen this item today" could not support good list memory performance. With this manipulation, after just 1 or 2 list presentations, all items become very familiar to the monkeys. If they are simply using familiarity judgments, they should respond "same" more often on "different" trials. Thus, under this condition, memory performance should become very poor, unless as in humans, these monkeys could remember that the test on a "different" trial does not match the current sample/listitem even though they may have seen it on the previous trial. Said otherwise, they would need to maintain memory of which items they have seen that day and in which list (i.e., prior list) they have seen them. Given that such strategies have been thought to depend on the integrity of the hippocampus, we would predict that removal of the hippocampus would render the subjects more susceptible to increases in proactive interference from previous list items and impair memory performance.

4.1. Method

Cuba was the only monkey that was available to be tested for an across list interference effect on list memory performance. For this testing, the stimulus set size was progressively reduced from 144, 32, 16 and then 8 stimuli with tests at the shortest delay of 0 s and the longest delay of 30 s and different sets of stimuli at the two delays. Finally, the effect of interference on serial position functions was assessed using a stimulus set of 8 items with all 6 probe delays (0 s, 1 s, 2 s, 10 s, 20 s and 30 s) in the block design previously described. In this phase of testing, one set of 8 stimuli was used for delays of 0 s, 1 s, and 2 s, and a different set of 8 stimuli was used for delays of 10 s, 20 s, and 30 s, so that tests in the first block of a daily test session would not adversely affect (buildup of proactive interference) and confound tests in the second block.

4.2. Results

As shown in Fig. 6, it is remarkable that Cuba maintained excellent performance even when the set size was decreased from 144 items to 8 items. Thus, despite the high level of proactive interference and overall high level of familiarity created by the use of a small set of 8 items, this animal nevertheless maintained a strong recency effect at the short delay of 0 s and a strong primacy effect at the long delay of 30 s. In addition, memory performance remained high in this animal when the six delays were blocked. As shown in Fig. 7, Cuba's performance shows a strong recency effect at the 1 s delay, and, as the delays increased, memory for the first items of the list became stronger (primacy effect) and progressively inhibited memory for the last items.

5. Discussion

The main findings of this study indicate that contrary to most memory models of the role of the hippocampus, selective damage to the hippocampus had only minor effects on list memory accuracy and dynamic time changes of the serial position function. Most memory models would predict: loss of primacy (first item) memory, loss of accuracy with delay, and loss in the ability to combat proactive interference. Contrary to those predictions, primacy memory actually increased for two monkeys (Gracie and Slim) at shorter delays and recency memory increased for a third monkey (Cuba) at longer delays, accuracy was maintained at long delays including delays up to 240 s (8 times longer than any pre-surgery testing), and accuracy was maintained despite high levels of proactive interference. Implication of these findings is discussed in the following sections.





Fig. 4. Serial position functions curves at the end of list-learning training for Gracie, Slim, and Cuba. For each animal, top graphs illustrate pre-surgery performance and bottom graphs illustrate post-surgery performance. Data points are mean performance for four-item serial lists at different probe delays (retention intervals), the interval between the last list item and the probe test item. The bar shown for each serial-position function is the average standard error of the mean for the four serial positions ("same" trials). Open triangles show performance on "different" trials where the probe item matched none of the four list items.



Fig. 5. Cuba and Slim post-surgical serial position functions with delays increased to 60 s, 120 s, and 240 s. Error bars are the average standard error of the mean for the four serial positions of each function.

5.1. Hippocampus and serial list memory performance

Two monkeys, Gracie and Cuba, were shown to perform the S/D task to high levels of proficiency in only 64 trials (i.e., no loss of learning) following hippocampal lesions. These monkeys also showed full abstract-concept learning where transfer was equivalent to baseline performance. A third monkey, Slim required some re-learning (1636 trials), but there was a huge savings relative to the original acquisition (30,600 trials) and he also eventually achieved accurate performance and full abstract-concept learning. All three monkeys also demonstrated good performance on the serial list memory task at 0-30s delays, with strong recency and primacy effects. The lack of effects of hippocampal lesions on the serial position functions differs from the absence of primacy effect that has been previously reported in patients with medial temporal lobe damage (e.g., Baddeley and Warrington, 1970; Hermann et al., 1996; Hopkins et al., 1995; Hopkins and Kesner, 1995) as well as rodents with selective hippocampal damage (Kesner, 1998; Kesner and Novak, 1982). The difference in results most likely stems from several differences between these previous studies and the one reported here.

First, differences between the human and monkey data could have resulted from varying extent of medial temporal lobe tissue included in the hippocampal lesions. In the earlier human studies, damage to the hippocampus was not selective and extended to the surrounding cortical areas that may have had greater impact on memory performance. Alternatively, one could argue that in two cases of the present study the hippocampal damage was not complete and the spared hippocampal tissue may have allowed for good performance. But this latter possibility seems unlikely given that in one case (Cuba) the hippocampal damage was selective, bilateral, and complete and nevertheless Cuba's recognition memory accuracy remained virtually unchanged from his pre-surgery levels. Another factor that could be used to explain the different outcomes between the human and monkey results was a difference in list lengths. Longer lists have been used in many human studies than the list of 4 items used here with the monkeys. Longer lists do emphasize the primacy effect more than shorter lists (Murdock, 1962). Nevertheless, there are a number of mitigating factors that bear upon this issue. For example, none of the human studies with longer lists systematically varied the retention interval. And varying the retention interval does emphasize the visual memory

primacy effect. Moreover, by varying the retention interval the return-to-primacy or Law of Primacy becomes apparent (see Tulving, 2008; Wright, 2013, this volume). Moreover, primacy effects have also been shown for humans with short 4-item lists of kaleidoscope patterns along with similar dynamic SPF changes with delay (e.g., Wright et al., 1985).

Second, differences between the rodent and monkey data cannot be accounted by differences in the extent of hippocampal lesion. The rodents, however, were trained in a radial-arm maze which tends to favor memory for spatial locations instead of memory for items. Given the well-known role of the hippocampus in spatial memory (for review see O'Keefe and Nadel, 1978; Kessels et al., 2001; King et al., 2004), differences between the outcomes of rodent studies and the monkey study of this article could be due to the difference between spatial memory ("where" memory) and item memory ("what" memory) of the present study.

Additionally, the intact performance of the three monkeys on the list memory task actually is supported by other findings of unimpaired recognition memory performance by monkeys, rodents, and amnesic patients. Monkeys with selective hippocampal lesions have demonstrated intact recognition memory in the delayed nonmatching-to-sample(DNMTS) task, even with substantial memory delays and long lists of sample stimuli. Monkeys with selective hippocampal lesions have demonstrated intact recognition memory in memory span tasks (Heuer and Bachevalier, 2011; Murray and Mishkin, 1998; Nemanic et al., 2004; but see also Beason-Held et al., 1999; Zola et al., 2000). Similarly, DNMTS recognition performance in rodents is also intact after selective hippocampal damage, although other studies report partial impairment at long delays or long lists (Clark et al., 2000, 2001; Mumby, 2001; Steckler et al., 1998). Intact recognition memory was also found in several cases of amnesic patients with damage to the hippocampal formation, specifically in those tested with a ves/no recognition task, a task similar to the S/D task used in the current study (Adlam et al., 2009; Mayes et al., 2002; Vargha-Khadem et al., 1997; but see also Cipolotti et al., 2001; Manns et al., 2003).

5.2. Hippocampus and within-list interference

Given that one well-accepted role of the hippocampus in memory processes is to combat interference among items of a list (Eichenbaum and Buckingham, 1990; Shapiro and Olton, 1994), we



Fig. 6. Serial position functions for Cuba with the set size of items progressively reduced from 144 to 32, 16 and finally 8 items. Delays used were 0 s and 30 s. Error bars are the average standard error of the mean for the four serial positions of each function.

predicted two possible outcomes following hippocampal lesions. First, retroactive interference of the last item upon the first item of the list would likely increase such that the primacy effect would be delayed or eliminated, as had been shown in previous studies with humans (Baddeley and Warrington, 1970; Hermann et al., 1996; Hopkins et al., 1995; Hopkins and Kesner, 1995) and rodents (Kesner, 1998; Kesner and Novak, 1982). Alternatively, under trial-unique conditions where the first list items exert proactive interference on memory for the last list items, removal of the hippocampus may increase this within-list proactive interference and actually delay or eliminate the recency effect. The data showed that, after selective hippocampal lesions, the shapes of some serial position curves were slightly altered for all 3 animals, but contrary to our expectations neither the primacy effect nor the recency effect was delayed or abolished (see Fig. 4).

Extensive pre-surgical training may have allowed the animals to rely on their well-learned reference memory of the task, which in turn may have provided the necessary basis for good working memory (remembering items of the current list) and their robust retrieval memory.

5.3. Hippocampus and across-list interference

Several studies have provided evidence that not all recognition memory tasks are sensitive to hippocampal lesions (Holdstock et al., 2002; Nemanic et al., 2004; Reed et al., 1997) and more particularly that memory performance on the DNMTS and yes/no recognition tasks may be supported by familiarity judgments that do not require the contribution of the hippocampus (for review see Aggleton and Brown, 2006; Diana et al., 2007; Eichenbaum et al., 2007). This proposal was specifically tested in the second experiment that minimized the effectiveness with which animals could use a familiarity strategy by increasing item repetition across trials and proactive interference. We did this by reducing the stimulus set size from 432 items (trial unique per daily session) to 8 items so that items would be repeated every few trials and proactive interference would accumulate over the block of 32 trials tested at a particular delay. Hence, if monkeys were simply using familiarity judgments, then they should respond "same" more often and increase their "same" bias as the trial block wore on. But as shown in Figs. 6 and 7, there is no evidence supporting a familiarity judgment strategy for the animal with hippocampal damage tested in this experiment. Contrary to the familiary strategy, this animal maintained high levels of memory performance even at the longest delays tested. This result suggests that this animal was using a strategy (or strategies) other than familiarity judgments to solve the task and recollect whether a memory item was in the current list or in some previous list. Whether they shifted their strategy from familiarity with session-unique stimuli to a recollective memory processes with repeated stimuli or were using a more recollective process all along is an open question. Other studies have shown that monkeys trained and tested with repeated stimuli and high proactive interference do gradually improve their performance (see for review Wright, 2007). Either way, the extensive training this monkey received prior to surgery was likely instrumental in its post-surgery accurate performance.

Lastly, the good memory performance of Cuba that had nearly complete, bilateral, damage to the hippocampus is intriguing and needs further investigation because it contradicts several theories that have been advanced to support the role the hippocampus in episodic memory. For example, two proposals have mainly driven research on the role of the hippocampus in recognition memory. The first is the one-process theory (Squire, 1994; Squire et al., 2007) that suggests that familiarity and recollection can both be supported by the hippocampus when memory strength of signals is high. The second is the two-process theory proposing that the hippocampus is important for recognition memory for events of episodes (what, where and when) but not for familiarity judgments (Eichenbaum et al., 2007). However, neither theory would predict that Cuba without a functional hippocampus (bilaterally), could accurately perform the list-memory task under high proactive interference with only a familiarity strategy. The findings presented in this article will need to be followed up and replicated in a larger number of animals, perhaps with differing amounts of pre-operative training in the serial list memory task. Nevertheless, if confirmed these findings may change current conceptions of the role of the hippocampus in recognition memory and further



Fig. 7. Serial position functions across retention delays with a set size of 8-items for Cuba. Error bars are the average standard error of the mean for the four serial positions of each function.

our understanding underlying hippocampal-dependent processes of relational memory, familiarity, episodic memory.

Acknowledgments

This research was supported by National Institutes of Health Grant MH-58846 to J.B., and MH072616 and MH091038 to A.A.W. The authors would like to thank Jacquelyne J. Rivera for her unwavering dedication in conducting the research. The authors wish to express their appreciation to Christopher S. Smith, Terry L. Blasdel, Peggy Bek, and Tammy L. Humbird for their expertise and valuable help in the pre- and post-surgical care of the monkeys, Roger E. Price and Belinda Rivera for the care and handling of the animals during the MR imaging procedures, Edward F. Jackson for his expert assistance in neuroimaging scanning techniques, and Maria C. Alvarado for help in the neuroimaging and surgical procedures. Writing of this manuscript was also supported in part by the Yerkes Base Grant NIH RR00165, and the Center for Behavioral Neuroscience grant NSF IBN-9876754.

References

- Adlam, A.L., Malloy, M., Mishkin, M., Vargha-Khadem, F., 2009. Dissociation between recognition and recall in developmental amnesia. Neuropsychologia 47 (11), 2207-2210.
- Aggleton, J.P., Brown, M.W., 2006. Interleaving brain systems for episodic and recognition memory. Trends Cogn. Sci. 10, 455-463.
- Aggleton, J.P., Shaw, C., 1996. Amnesia and recognition memory: a re-analysis of psychometric data. Neuropsychologia 34(1), 51-62.
- Bachevalier, J., Mishkin, M., 1986. Visual recognition memory impairment follows ventromedial but not dorsolateral prefrontal lesions in monkeys. Behav. Brain Res. 20, 249-261.
- Baddeley, A.D., Warrington, E.K., 1970. Amnesia and the distinction between short and long-term memory. J. Verbal Learn. Verbal Behav. 9, 176–189. Barker, G.R.I., Warburton, E.C., 2011. When is the hippocampus involved in recog-
- nition memory? J. Neurosci. 31, 10721-10731.
- Bastin, C., Linden, M., Charnallet, A., Denby, C., Montaldi, D., Roberts, N., Andrew, M., 2004. Dissociation between recall and recognition memory performance in an amnesic patient with hippocampal damage following carbon monoxide poisoning. Neurocase 10, 330-344.
- Baxter, M.G., Murray, E.A., 2001. Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. Hippocampus 11 (1), 61-71.
- Beason-Held, L.L., Rosene, D.L., Killiany, R.J., Moss, M.B., 1999. Hippocampal formation lesions produce memory impairment in the rhesus monkey. Hippocampus 9 (5), 562-574.
- Belmont, J.M., Butterfield, E.C., 1971. What the development of short-term memory is? Hum. Dev. 14 (4), 236-248.
- Bhatt, R.S., Wright, A.A., 1992. Concept leaning by monkeys with video picture images and a touch screen. J. Exp. Anal. Behav. 57, 219-225.
- Bowles, B., Crupi, C., Mirsattari, S.M., Pigott, S., Parrent, A.G., Pruessner, J.C., Yonelinas, A.P., Kohler, S., 2007. Impaired familiarity with preserved recollection after anterior temporal lobe resection that spares hippocampus. Proc. Natl. Acad. Sci. U.S.A. 104, 16382-16387.
- Brown, M., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Prog. Neurobiol. 55. 149–189.
- Cipolotti, L., Shallice, T., Chan, D., Fox, N., Scahill, R., Harrison, G., Stevens, J., Rudge, P., 2001. Long-term retrograde amnesia. the crucial role of the hippocampus. Neuropsychologia 39 (2), 151-172.

- Clark, R.E., West, A.N., Zola, S.M., Squire, L.R., 2001. Rats with lesions of the hippocampus are impaired on the delayed nonmatching-to-sample task. Hippocampus 11, 176-186
- Clark, R.E., Zola, S.M., Squire, L.R., 2000. Impaired recognition memory in rats after damage to the hippocampus. J. Neurosci. 20, 8853-8860.
- Crowder, R.G., 1986. Auditory and temporal factors in the modality effect. J. Exp. Psychol. Learn. Mem. Cogn. 12 (2), 268-278.
- Crowder, R.G., Morton, J., 1969. Precategorical acoustic storage (PAS). Percept. Psychophys. 5 (6), 365-373.
- Damasio, A.R., Eslinger, P.J., Damasio, H., Van Hoesen, G.W., Cornell, S., 1985. Multimodal amnesic syndrome following bilateral temporal and basal forebrain damage. Arch. Neurol. 42 (3), 252-259
- Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn. Sci. 11 (9), 379-386.
- Eichenbaum, H., Buckingham, J., 1990. Studies on hippocampal processing: experiment, theory and model. In: Gabriel, M., Moore, J. (Eds.), Learning and Computational Neuroscience: Foundation of Adaptive Networks. MIT, Cambridge, MA, pp. 171-231.
- Eichenbaum, H., Mathews, P., Cohen, N.J., 1989. Further studies of hippocampal representation during odor discrimination learning. Behav. Neurosci. 103 (6), 1207-1216.
- Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and recognition memory. Annu. Rev. Neurosci. 30, 123-152.
- Ennaceur, A., Neave, N., Aggleton, J.P., 1996. Spontaneous object recognition and object location memory in rats: the effects of lesions in the cingulate cortices, the medial prefrontal cortex, the cingulum and the fornix. Exp. Brain Res. 113, 509-519.
- Fahy, F., Riches, I., Brown, M., 1993. Neuronal activity related to visual recognition memory, long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Exp. Brain Res. 96, 457-472.
- Freed, D.M., Corkin, S., Cohen, N.J., 1987. Forgetting in H.M.: a second look. Neuropsychologia 25, 461–471.
- Gaffan, D., Murray, E.A., 1992. Monkeys (Macaca fascicularis) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertribal intervals and fail matching to sample despite double sample presentations. Behav, Neurosci, 106, 30-38.
- Gardiner, J.M., 1974. Levels of processing in word recognition and subsequent free recall. J. Exp. Psychol. 102 (1), 101-105.
- Glanzer, M., Cunitz, A.R., 1966. Two storage mechanisms in free recall. J. Verbal Learn. Verbal Behav. 5, 351–360. Goldman-Rakic, P.S., 1987. Circuitry of the frontal association cortex and its rele-
- vance to dementia, Arch. Gerontol, Geriatr. 6 (3), 299-309.
- Hermann, B.P., Seidenberg, M., Wyler, A., Davies, K., Christeson, J., Moran, M., Stroup, E., 1996. The effects of human hippocampal resection on the serial position curve. Cortex 32 (2), 323-334.
- Heuer, E., Bachevalier, J., 2011. Effects of selective neonatal hippocampal lesions on tests of object and spatial recognition memory in monkeys. Behav. Neurosci. 125, 137-149.
- Hodos, W., Bobko, P.A., 1984. A weighted index of bilateral brain lesions. J. Neurosci. Methods 12, 43-47.
- Holdstock, J.S., Mayes, A.R., Gong, Q.Y., Roberts, N., Kapur, N., 2005. Item recognition is less impaired than recall and associative recognition in a patient with selective hippocampal damage. Hippocampus 15, 203-215.
- Holdstock, J.S., Mayes, A.R., Roberts, N., Cezayirli, E., Isaac, C.L., O'Reilly, R.C., Norman, K.A., 2002. Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? Hippocampus 12, 341-351.
- Hopkins, R.O., Kesner, R.P., 1995. Item and order recognition memory in subjects with hypoxic brain injury. Brain Cogn. 27, 180–201.
- Hopkins, R.O., Kesner, R.P., Golsdtein, M., 1995. Memory for novel and familiar spatial and linguistic temporal distance information in hypoxic subjects. J. Int. Neuropsychol. Soc. 1 (5), 454-468.
- Jones, B.M., 1973. Memory impairment on the ascending and descending limbs of the blood alcohol curve. J. Abnorm. Psychol. 82 (1), 24-32.
- Katz, J.S., Wright, A.A., Bachevalier, J., 2002. Mechanisms of same/different abstractconsept learning by rhesus monkeys (Macaca mulatta). J. Exp. Psychol. Anim. Behav. Process 28 (4), 358-368.

Kesner, R.P., 1985. Correspondence between humans and animals in coding of temporal attributes: role of hippocampus and prefrontal cortex. Ann. N.Y. Acad. Sci. 444, 122-36.

- Kesner, R.P., 1998. Neural mediation of memory for time: role of the hippocampus and medial prefrontal cortex. Psychon. Bull. Rev. 5, 585–596.
- Kesner, R.P., Novak, J.M., 1982. Serial position curve in rats: role of the dorsal Hippocampus. Science 218, 173–174.
- Kessels, R.P.C., de Haan, E.H.F., Kappelle, L.J., Postma, A., 2001. Varieties of human spatial memory: a meta-analysis on the effects of hippocampal lesions. Brain Res. Rev. 35, 295–303.
- King, J.A., Trinkler, I., Hartley, T., Vargha-Khadem, F., Burgess, N., 2004. The hippocampal role in spatial memory and the familiarity-recollection distinction: a case study. Neuropsychology 18 (3), 405–417.
- Kolb, B., Buhrmann, K., McDonald, R., Sutherland, R.J., 1994. Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory. Cereb. Cortex 4, 664–680.
- Málková, L., Lex, C.K., Mishkin, M., Saunders, R.C., 2001. MRI-based evaluation of locus and extent of neurotoxic lesions in monkeys. Hippocampus 11, 361–370.
- Manns, J.R., Hopkins, R.O., Reed, J.M., Kitchener, E.G., Squire, L.R., 2003. Recognition memory and the human hippocampus. Neuron 37 (1), 171–180.
- Mayes, A.R., Holdstock, J.S., Isaac, C.L., Hunkin, N.M., Roberts, N., 2003. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. Hippocampus 12, 325–340.
- Mayes, A.R., Isaac, C.L., Holdstock, J.S., Cariga, P., Gummer, A., Roberts, N., 2002. Relative sparing of item recognition memory in a patient with adult-onset damage limited to the hippocampus. Hippocampus 12, 325–340.
- Meunier, M., Bachevalier, J., Mishkin, M., Murray, E., 1993. Effects on visual recognition of combined and separate abaltions of the entorhinal andperirhinal cortex in rhesus monkeys. J. Neurosci. 13, 5418–5432.
- Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J. Neurosci. 16, 5154–5167.
- Mishkin, M., Vargha-Khadem, F., Gadian, D.G., 1998. Amnesia and the organization of the hippocampal system. Hippocampus 8, 212–216.
- Mumby, D.G., 2001. Perspectives on object-recognition memory following hippocampal damage: lessons from studies in rats. Behav. Brain Res. 127 (1/2), 159–181.
- Murdock Jr., B.B., 1962. The serial position effect in free recall. J. Exp. Psychol. 64, 482–488.
- Murdock Jr., B.B., 1966. Visual and auditory stores in short-term memory. Q. J. Exp. Psychol. 18 (3), 206–211.
- Murray, E., Bussey, T., 1999. Perceptual-mnemonic functions of the perirhinal cortex. Trends Cogn. Sci. 3, 142–151.
- Murray, E.A., Mishkin, M., 1998. Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. J. Neurosci. 18, 6568–6582.
- Nemanic, S., Alvarado, M.C., Bachevalier, J., 2004. The hippocampal/parahippocampal regions and recognition memory: insights from visual paired comparison versus object-delayed nonmatching in monkeys. J. Neurosci. 24, 2013–2026.
- Nemanic, S., Alvarado, M.C., Price, R.E., Jackson, E.F., Bachevalier, J., 2002. Assessment of locus and extent of neurotoxic lesions in monkeys using neuroimaging techniques: a replication. J. Neurosci. Methods 121, 199–209.
- Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110, 611–646.
- O'Keefe, J., Nadel, L., 1978. The Hippocampus as a Cognitive Map. Oxford University Press, Oxford, UK.
- Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, T., Hamalainen, A., Soininen, H., Aronen, H.J., 2004. Visual representation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe subareas in humans. Eur. J. Neurosci. 19, 1939–1949.
- Postman, L., Phillips, L., 1965. Short-term temporal changes in free-recall. Q. J. Exp. Psychol. 17, 132–138.
- Reed, J.M., Hamann, S.B., Stefanacci, L., Squire, L.R., 1997. When amnesic patients perform well on recognition memory tests? Behav. Neurosci. 111, 1163–1170.
- Roediger III, H.L., Crowder, R.G., 1976. A serial position effect in recall of United States presidents. Bull. Psychon. Soc. 8, 275–278.

Rudy, J.W., Sutherland, K.J., 1989. The hippocampal formation is necessary for rats to learn and remember configural discriminations. Behav. Brain Res. 34, 97–109.

- Rudy, J.W., Sutherland, K.J., 1992. Configural and elemental associations and the memory coherence problem. J. Cogn. Neurosci. 4 (3), 208–216.
- Rugg, M.D., Yonelinas, A.P., 2003. Human recognition memory: a cognitive neuroscience perspective. Trends Cogn. Sci. 7, 313–319.
- Saffran, E.M., Marin, O.S., 1975. Immediate memory for word lists and sentences in a patient with deficient auditory short-term memory. Brain Lang. 2 (4), 420–433.
- Shapiro, M.L., Olton, D.S., 1994. Hippocampal function and interference. In: Schacter, D.L., Tulving, E. (Eds.), Memory Systems. The MIT Press, Cambridge, MA, pp. 87–117.
- Simons, J.S., Spiers, H.J., 2003. Prefrontal and medial temporal lobe interactions in long-term memory. Nat. Rev. Neurosci. 4, 637–648.
- Skinner, E.I., Fernandes, M.A., 2007. Neural correlates of recollection and familiarity: a review of neuroimaging and patients data. Neuropsychologia 45, 2163–2179.
- Squire, L.R., 1994. Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. In: Schacter, D.L., Tulving, E. (Eds.), Memory Systems. The MIT Press, Cambridge, MA, pp. 203–231.
- Squire, L.R., Wixted, J.T., Clark, R.E., 2007. Recognition memory and the medial temporal lobe: a new perspective. Nat. Rev. Neurosci. 8, 872–883.
- Stark, C.E., Bayley, P.J., Squire, L.R., 2002. Recognition memory for single items and for associations is similarly impaired following damage limited to the hippocampal region. Learn. Mem. 9, 238–242.
- Steckler, T., Drinkenburg, W.H., Saghal, A., Aggleton, J.P., 1998. Recognition memory in rats. II. Neuroanatomical substrates. Prog. Neurobiol. 54, 313–332.
- Sutherland, R.J., Rudy, J.W., 1989. Configural association theory: the role of the hippocampal formation in learning, memory, and amnesia. Psychobiology 17 (2), 129–144.
- Suzuki, W., Zola-Morgan, S., Squire, L., Amaral, D., 1993. Lesions of the perirhinal and parahippocampal cortices in monkeys produce long-lasting memory impairment in visual and tactual modalities. J. Neurosci. 13, 2430–2451.
- Tulving, E., 2008. On the law of primacy. In: Gluck, M.A., Anderson, J.R., Kosslyn, S.M. (Eds.), Memory and Mind. Erlbaum, New York, NY, pp. 31–48.
- Vargha-Khadem, F., Gadian, D.G., Watkins, K.E., Connelly, A., Van Paesschen, W., Mishkin, M., 1997. Differential effects of early hippocampal pathology on episodic and semantic memory. Science 277 (53234), 376–380.
- von Bonin, G., Bailey, P., 1947. The Neocortex of Macaca mulatta. University of Illinois Press, Urbana, IL.
- Wais, P.E., 2008. fMRI signals associated with memory strength in the medial temporal lobes: a meta-analysis. Neuropsychologia 46, 3185–3196.
- Warrington, E.K., Logue, V., Pratt, R.T., 1971. The anatomical localisation of selective impairment of auditory verbal short-term memory. Neuropsychologia 9 (4), 377–387.
- Warrington, E.K., Shallice, T., 1984. Category specific semantic impairments. Brain 107 (3), 829–854.
- Watkins, M.J., Watkins, O.C., 1974. Processing of recency items for free recall. J. Exp. Psychol. 102 (3), 488–493.
- Weiskrantz, L., 1987. Neuroanatomy of memory and amnesia: a case for multiple memory systems. Hum. Neurobiol. 6 (2), 93–105.
- Wixted, J.T., Mickes, L., Squire, L.R., 2010. Measuring recollection and familiarity in the medial temporal lobe. Hippocampus 20 (11), 1195–1205.
- Wright, A.A., 2007. An experimental analysis of memory processing. J. Exp. Anal. Behav. 88, 405–433.
- Wright, A.A., Katz, J.S., 2006. Mechanisms of same/different concept learning in primates and avians. Behav. Process. 72, 234–254.
- Wright, A.A., Rivera, J.J., Katz, J.S., Bachevalier, J., 2003. Abstract-concept learning and list-memory processing by capuchin and rhesus monkeys. J. Exp. Psychol. Anim. Behav. Process. 29, 184–198.
- Wright, A.A., Santiago, H.C., Sands, S.F., Kendrick, D.F., Cook, R.G., 1985. Memory processing of serial lists by pigeons, monkeys, and people. Science 229 (4710), 287–289.
- Wright, A.A., 2013. Functional Relationships for Investigating Cognitive Processes. Behav. Process. 93, 4–24.
- Xiang, J.Z., Brown, M.W., 2004. Neuronal responses related to long-term recognition memory processes in prefrontal cortex. Neuron 42, 817–829.
- Yonelinas, A.P., Parks, C.M., 2007. Receiver operating characteristics (ROCs) in recognition memory: a review. Psychol. Bull. 133 (5), 800–832.
- Zola, S.M., Squire, L.R., Teng, E., Stefanacci, L., Buffalo, E.A., Clark, S.K., 2000. Impaired recognition memory in monkeys after damage limited to the hippocampal region. J. Neurosci. 20, 451–463.