

Game-like tasks for comparative research: Leveling the playing field

DAVID A. WASHBURN

Georgia State University, Atlanta, Georgia
and Morris Brown College Center of Excellence for Research on Training, Atlanta, Georgia

and

JONATHAN P. GULLEDGE

Georgia State University, Atlanta, Georgia

Game-like computer tasks offer many benefits for psychological research. In this paper, the usefulness of such tasks to bridge population differences (e.g., age, intelligence, species) is discussed and illustrated. A task called ALVIN was used to assess humans' and monkeys' working memory for sequences of colors with or without tones. Humans repeated longer lists than did the monkeys, and only humans benefited when the visual stimuli were accompanied by auditory cues. However, the monkeys did recall sequences at levels comparable to those reported elsewhere for children. Comparison of similarities and differences between the species is possible because the two groups were tested with exactly the same game-like paradigm.

Modern computer games provide an undeniable source of entertainment, challenge, diversion, education, and activity. They attract participation by individuals across demographic boundaries of age, gender, ethnic origin, and educational status (and even, we shall see, species). Games permit subjects to experience novel events and challenges, to explore new worlds, to engage in extraordinary activities, and to destroy or even die without real consequences. Games also frequently require a range of complex behaviors reflecting processes such as attention, learning, memory, problem solving, planning, mental rotation, and psychomotor functioning. This notwithstanding, these psychological constructs have seldom been systematically examined using computer-game paradigms.

Computer games hold the promise of numerous benefits for psychological research. They elicit motivated performance and provide context and continuity for responding. The challenge for the investigator is to transform "computer games"—without losing the aspects that make them engaging and useful—into "computerized tasks" or tests that permit manipulation and control of independent variables, collection of dependent measures,

and the other necessary aspects of experimental design. This, of course, is necessary so that psychologists can *do* good science with game-like paradigms; additionally, the research must be *perceived* as good science rather than as "wasting time and resources playing games." It seems likely that researchers who might otherwise be interested in the behavioral phenomena produced in computer games are dissuaded from their use by the fear that their research will appear sophomoric or trivial.

To combat these hazards, it would seem to be advisable to avoid the term *games* in research, unless of course one is studying "the psychology of computer-game playing." The goal here is neither to conceal nor to confuse the issues using terminology but rather to recognize that activities selected for scientific tests are chosen, if not designed, to address specific empirical issues. For some game-like tests, the term *simulation* has been appropriately applied. Simulations permit users to pretend to engage in an activity (e.g., flying a plane) rather than performing the activity in reality. Quite in contrast with *games*, the term *simulation* carries with it a connotation of complexity and sophistication. However, many other computer games can be used for psychological research (e.g., chess) but are not appropriately considered to be simulations (i.e., one typically cannot *be* the chess pieces). Consequently, the use of *games* to describe recreational software, and the use of computer *tasks* to denote programs that are used for scientific purposes, is recommended even in cases in which no other difference separates the two.

For example, Rumbaugh and his colleagues have developed a battery of tasks designed to assess basic cognitive and psychomotor functioning in nonhuman primates (Rumbaugh, Richardson, Washburn, Savage-Rumbaugh,

This research was supported in part by Grant NAG2-438 from the National Aeronautics and Space Administration to Georgia State University, and by Army Research Office Grant DAAL03-92-G-0382 to the Morris Brown College Center of Excellence for Research on Training. The authors thank Duane M. Rumbaugh and the students in Introduction to Learning for their contributions to this research, and Herb Terrace for his comments on portions of the Discussion. Correspondence concerning this article or requests for the ALVIN task should be addressed to D. A. Washburn, Sonny Carter Life Sciences Laboratory, Department of Psychology, Georgia State University, Atlanta, GA, 30303 (e-mail: psydaw@gsuvm1.gsu.edu).

& Hopkins, 1989). Each of these tasks requires subjects (monkeys or apes) to respond to computer-graphic stimuli by manipulating a joystick. The tasks could be considered to be games. They look like simple computer games and are highly engaging for the animals. Rhesus monkeys work on the tasks (note that subjects "work on tasks" vs. "play the games") for many hours each day, even when they have abundant food from other sources (Washburn & Rumbaugh, 1992). However, each task represents a carefully designed, automated version of a classic testing paradigm in cognitive (e.g., visual search, divided visual field) or comparative psychology (e.g., matching-to-sample, discrimination learning).

In the present paper, the use of these game-like computer tasks to bridge population boundaries will be illustrated. The present discussion will emphasize the utility of these tasks for spanning species barriers, but they are at least equally applicable to bridging age, gender, ethnic, or other differences. When the same tasks can be administered, without significant alteration, to representatives of different target populations—so that groups are playing on a level, or at least a common, field—unprecedented degrees of comparability and generality can be obtained.

The present investigation was designed to assess the ability of humans and rhesus monkeys to remember sequences for short periods of time. Typically, the tasks used to study working memory with human subjects are quite different from those used with nonhuman animals. The task for this study was inspired by the familiar handheld computer game called SIMON (trademark, Milton Bradley). In this game, random sequences of lights and tones are presented, each of which must be reproduced by the player. Following successful reproductions of the light+tone sequence, the series is augmented with a new stimulus, requiring the player to execute ever longer strings of responses. Errors cause a new sequence to be presented, beginning with only one light+tone. It was anticipated that such a task would be ideal for studying working memory for sequences, in that individuals could respond across trials to progressively more challenging memory sets without intervention or calibration of task demands by the experimenter.

METHOD

Subjects

We tested 50 human subjects (age range 18–38 years; 38 females, 12 males), who volunteered to participate in exchange for course credit. These students received instructions prior to beginning the experiment. Five rhesus monkeys (*Macaca mulatta*; 1 female, 4 males) were also tested. The monkeys were trained to manipulate a joystick to respond to computer-generated stimuli prior to this study (Rumbaugh et al., 1989). The monkeys were not deprived of food or water for testing, and they worked ad lib on the task.

Apparatus

All subjects were tested with the Language Research Center's Computerized Test System (Rumbaugh et al., 1989). This apparatus consisted of a 386-based computer, which was used to gener-

ate VGA graphics stimuli on a color monitor. The subjects responded to these stimuli by manipulating a standard analog joystick, which in turn controlled the movements of a cursor (a small "+" on the screen). Auditory feedback was presented through the computer's internal speaker or via an external speaker/amplifier. For the monkeys, successfully completed sequences were also followed by automatic delivery of fruit-flavored pellets.

The task used in this experiment was called ALVIN. Each problem began after the subject moved the cursor into a box located in the middle of the screen. Consequently, four colored rectangles were immediately presented on the black screen (a 12 × 3 cm blue rectangle across the top of the screen, a 3 × 12 cm magenta rectangle down the right border, a green rectangle at the bottom, and a red on the left). One rectangle, selected at random, was illuminated by outlining it in yellow and switching to a lighter shade of color (i.e., blue became cyan, red became pink, etc.). The rectangle was lighted for 0.5 sec. Each colored rectangle was associated with one of four tones (500, 1000, 1500, and 2000 Hz), so that illumination could be accompanied by a color-specific auditory stimulus.

Like the commercial game that inspired it, each problem of the ALVIN task initially presented one color+tone stimulus. Subsequently, the cursor was presented in the middle of the screen, and the subjects were required to bring the cursor into contact with the colored rectangle that had been illuminated. Auditory feedback was provided and, if the response was correct, a pellet was dispensed (for the monkeys). On the second trial of the problem, two stimuli were presented: the original color+tone and a second color+tone, randomly selected, with the constraint that no stimulus was presented twice consecutively. Each problem continued in this manner, with increasingly long sequences of responses required on each trial (note that each problem consisted of an increasing number of trials, and each trial consisted of an increasing number of responses). A problem ended with the first error, which resulted in a raucous buzz during a time-out period. The length of the time-out period was inversely proportional to the number of stimuli in the sequence at the time of the error (10 sec – number of stimuli).

For each trial, the computer automatically recorded each stimulus sequence and each response sequence, together with the response latency and overall response time for producing the list. Additionally, time and date markers and a list of all parameter settings were recorded in the ASCII-format data files.

Procedure

Because the monkeys had never seen a task like ALVIN, they received training on a version of the task in which rectangles of color appeared on the screen only as they became part of the sequence (that is, the color did not appear until it was part of the sequence). Thus, only the one target stimulus was on the screen at the beginning of a problem. Two stimuli were on the screen on the second trial of a problem, and so forth. Each monkey performed 10,000 of these training trials prior to testing on the ALVIN task as described above. Subsequently, the monkeys performed 2,000 trials each under the task conditions originally described. On each of these test problems, the computer randomly determined whether or not to present tones along with the visual stimuli (i.e., half of the problems presented colors + tones, and half presented colors only).

Each human subject was tested for 40 min. On half of the problems they received, randomly determined, tone stimuli were omitted so that only visual stimuli were presented.

RESULTS

The monkeys performed an average of 1,140 trials per day. Two of the monkeys failed to learn the task; that is, their accuracy levels were about 50% on the training trials and did not show improvement across training. Data from

these 2 monkeys were omitted from subsequent analyses in this study, as they continue to receive training. The other 3 animals did manifest learning across training, improving their mean accuracy from 49% in the initial 1,000 trials to 68% in the terminal block of training.

Separate analyses were conducted for each species. Humans were able to remember longer sequences than were rhesus monkeys (7.03 vs. 3.68 items, respectively). For humans, the mean list length had a range of 4.13–24.00 items, with a standard deviation of 2.2. The monkeys produced a range of 1.0–13.0 items, with a standard deviation of 0.5. For the human subjects, significantly longer lists were remembered with color+tone stimuli than with color stimuli only (7.44 items vs. 6.63 items, respectively) [$t(49) = 2.47, p < .05$]. In contrast, no significant difference was observed for the monkeys' data (3.68 color only items, 3.67 color+tone items; $p > .05$).

Performance by humans and monkeys was affected by list length, as seen in the top panel of Figure 1. The humans were generally more accurate than were the monkeys and, consequently, produced longer sequences of responses. On the other hand, when performance was analyzed as a function of retention interval (the elapsed time between presentation of a stimulus and the corresponding response), a different pattern was evident (Figure 1, bottom). The species performed accurately with delays up to 12–15 sec, beyond which performance was disrupted. Note that the slopes for forgetting were com-

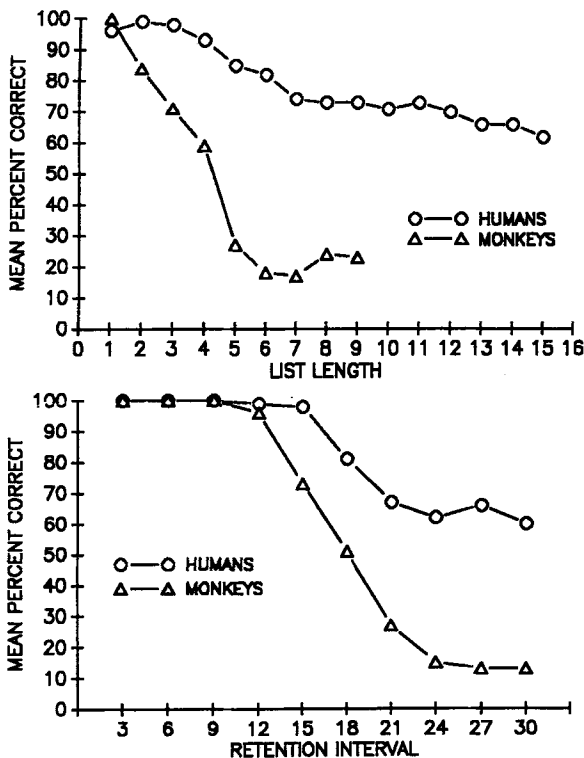


Figure 1. Mean percentage correct as a function of species and list length (top panel) and retention interval (bottom panel).

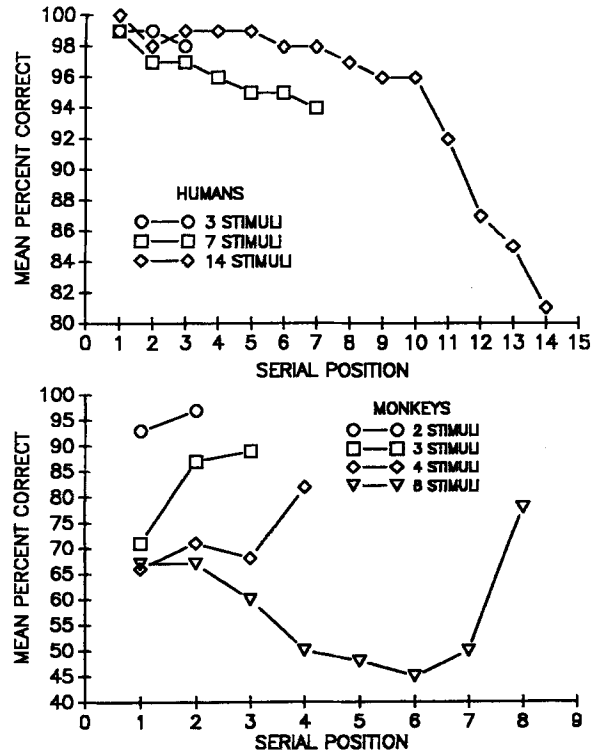


Figure 2. Mean percentage correct as a function of list length and serial position for human subject (top panel) and rhesus monkeys (bottom panel).

parable for the two species, but the asymptotes were markedly different.

Finally, the data were analyzed as a function of sequence length and serial position. Figure 2 reflects the results of these analyses for several list lengths with percentages of correct responses for each serial position.¹ Note that the human subjects tended to recall best those items at the beginning of the sequence, regardless of trial length (the data in Figure 2, top, represent serial position curves for the mean length, a lower anchor, and a long sequence of about twice the mean sequence length). In contrast, rhesus monkeys were significantly most likely to respond correctly on the last item in a sequence.

DISCUSSION

The human subjects and some rhesus monkeys were able to perform the ALVIN task, albeit with different levels of proficiency. Interestingly, the mean list length obtained for the monkeys in this experiment is comparable to the values reported for children's use of the commercial SIMON game (Moore & Staum, 1987). Those 5- to 7-year-old children repeated accurately an average of about 3.70 color+tone stimuli (vs. 3.68 for the monkeys). The human adults from the present study averaged 7.44 color+tone stimuli, which seems reasonable given the report of about 6-item recall for 10- and 11-year-old children (Moore & Staum, 1987). Thus, it appears that

the present data are within the range predicted from use of the commercial game.

That the monkeys were able to perform the task at these levels is noteworthy. Other researchers (e.g., D'Amato & Colombo, 1988; Terrace, 1993) have studied the capacity for monkeys to learn lists or sequences of arbitrary responses. The present paradigm is, of course, quite different from those studies in procedure and purpose. Terrace (1993) reviewed the simultaneous chaining paradigm, in which monkeys or pigeons learn to touch stimulus A, then stimuli A-B, then A-B-C, and so on to 4- or 6-item lists. In the simultaneous chaining paradigm, however, each new sequence consists of novel stimuli that are typically trained with a criterion-based, incremental procedure (although Terrace & McGonigle, 1994, reported that experienced monkeys can learn novel lists without this training). In the present task, the monkeys were required to reproduce arbitrary sequences of increasing length, each sequence using the exact same four color+tone stimuli, with only one presentation of each new sequence prior to responding.

Finally, the patterns of the serial position \times sequence length effects are particularly interesting in the present data. For human subjects, an unsurprising primacy effect was obtained. Subjects recalled best those items near the beginning of the sequence (which, of course, were presented and practiced most often). Relatively compromised performance at the end of the list is typical of the pattern humans produce in serial recall of verbal stimuli, particularly with delay intervals of the magnitude used here (see review by Glanzer, 1972).

For the monkeys, however, a different pattern is evident. Their bias for items at the end of the sequence resembles data normally obtained for human memory of visuospatial stimuli. Phillips and Christie (1977), for example, reported that human visual working memory (or the visuospatial sketch pad; Baddeley, 1990) has a capacity of one item, and that the last item in a list will be recalled or recognized best. This is exactly the pattern observed here. Of course, monkeys and other nonhuman animals can retain more than one stimulus—as, in fact, they demonstrated in the present study. Notwithstanding, the predisposition for remembering the terminal stimulus item, despite more practice on the preceding items, contrasts markedly with the humans' tendencies on this task.

Thus, it appears that the humans and the monkeys performed the ALVIN task with different levels of proficiency, and using markedly different strategies. The human

subjects appear to have made use of both stimulus modalities to retain the sequences as if they were auditory or articulatory information; the monkeys did not benefit from auditory stimuli, and they apparently maintained the sequences as visuospatial information. These comparisons were facilitated because both subject groups were tested with the exact same game-like task. Difficulty levels were calibrated for each subject automatically and, in fact, were used as an important dependent measure in the experiment. It is reasonable to conclude that many other computer games could be similarly useful tools for the scientific study of memory or other aspects of behavior and performance.

REFERENCES

- BADDELEY, A. D. (1990). *Human memory: Theory and practice*. Boston: Allyn & Bacon.
- D'AMATO, M. R., & COLOMBO, M. (1988). Representation of serial order in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 131-139.
- GLANZER, M. (1972). Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5, pp. 129-193). New York: Academic Press.
- MOORE, R. S., & STAUM, M. (1987). Effects of age and nationality on auditory/visual sequential memory of English and American children. *Bulletin of the Council for Research in Music Education*, **91**, 126-131.
- PHILLIPS, W. A., & CHRISTIE, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, **29**, 117-133.
- RUMBAUGH, D. M., RICHARDSON, W. K., WASHBURN, D. A., SAVAGE-RUMBAUGH, E. S., & HOPKINS, W. D. (1989). Rhesus monkeys (*Macaca mulatta*), video tasks, and implications for stimulus-response spatial contiguity. *Journal of Comparative Psychology*, **103**, 32-38.
- TERRACE, H. S. (1993). The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychological Science*, **4**, 162-169.
- TERRACE, H. S., & MCGONIGLE, B. M. (1994). Memory and representation of serial order by children, monkeys, and pigeons. *Current Directions in Psychological Science*, **3**, 180-185.
- WASHBURN, D. A., & RUMBAUGH, D. M. (1992). Investigations of rhesus monkey video-task performance: Evidence for enrichment. *Contemporary Topics in Laboratory Animal Science*, **31**, 6-10.

NOTE

- Note that retention curves in Figure 2 are higher than those in Figure 1. If an error occurred, for example, in the third position of a 3-item sequence, the entire trial was scored as "wrong" for Figure 1, but two thirds of the serial positions were scored "correct" for Figure 2.

(Manuscript received November 21, 1994;
revision accepted for publication January 31, 1995.)