*Genes, Brain and Behavior (2003) **2:** 40–55*

ISSN 1601-1848

The structure of individual differences in Heterogeneous Stock mice across problem types and motivational systems

C. Locurto*, E. Fortin and R. Sullivan

Department of Psychology, PO Box 75A, College of the Holy Cross, Worcester, MA 01610, USA
**Corresponding author: C. Locurto, Department of Psychology, PO Box 75A, College of the Holy Cross, Worcester, MA 01610, USA. E-mail: clocurto@holycross.edu*

Sixty Heterogeneous Stock (HS) mice received a battery of six problem-solving tasks and three control procedures. The problem-solving tasks included Hebb-Williams, a place learning task conducted in a plus maze, radial maze, a working memory test following the radial maze, a set of detour problems and a visual non-matching to sample task. The control procedures consisted of land and water activity measures and a light-dark test. The correlation matrix derived from these tasks did not exhibit positive manifold, that is, positive correlations across all problem-solving tasks. Principal components analysis reduced the correlation matrix to four components with eigenvalues exceeding 1.0. Instead of the general factor solution common in the study of human problem-solving, this component structure appeared more congenial to a modular interpretation, with the four components each explaining approximately the same magnitude of matrix variance.

Keywords: HS, individual differences, mice, motivation, problem-solving

Received 27 August 2002, revised 11 October 2002, accepted for publication 31 October 2002

One of the most contentious issues in the study of human nature derives from one of the most well documented findings in all of behavioral science. The finding is that if a group of subjects is given a battery of problem-solving tasks, the result is a matrix of positive correlations across these tasks (see Brody 1992; MacKintosh 1998, for reviews). That matrix, called positive manifold, means that individuals tend to retain their rank-ordering across tasks. This finding is contentious because it has led to suggestions that a single process or small set of processes, referred to as a general factor or *g*, may inform performance on all cognitively demanding tasks. It is a short theoretical step from there to suggest that individuals may be rank-ordered according to their expression of *g*, and, hence, that human differences in cognition are a matter of differences on this single dimension.

The documentation for positive manifold is, to say the least, extensive. It was first described by Spearman (1904) and has since become a standard finding across numerous test batteries (see Jensen 1998, for the most comprehensive recent summary of evidence). While the documentation for *g* theory at the human level is impressive, it has always lacked a comparative foundation that might help in understanding the evolutionary origins of *g*, or a comparable animal model that might facilitate exploration of the underlying neurobiology of *g* (Eysenck 1987; Humphreys 1987; Locurto 1997). The failure to develop a comparative basis to understand *g* is surprising given that so many aspects of human cognitive functioning have long been of central interest to students of comparative cognition (Hulse 1993; Terrace 1993; Weiskrantz 1985). This rather striking difference in the human and animal literatures derives in part from the lack of an individual difference research tradition in animal cognition, as well as from the failure of several early attempts to find an animal analog of *g* (see Locurto 1997, for a review).

Our first study of individual differences in mouse problem-solving (Locurto & Scanlon 1998) used two strains, the second filial generation cross between C57BL/6 and DBA/2Js inbred strains and a CD−1 outbred albino strain. We ran these strains through a water escape problem-solving battery consisting of Hebb-Williams, a place learning task conducted in a plus maze, spatial and visual reversal discriminations and Morris maze. The battery also included activity control procedures. Results indicated a positive manifold for both strains, for error and latency measures. The first principal component, which expresses the strength of *g* given positive manifold, accounted for between 28 and 61% of matrix variance depending on the strain and dependent measure. In contrast, human problem-solving batteries average between 40 and 60% (Jensen 1987, 1998).

There are three other contemporary research programs that have examined the structure of individual differences in animal problem-solving. Thompson and colleagues developed a battery for rats that contained a diverse array of problems (Thompson *et al.* 1990). Their battery included two detour problems, a water maze problem, spatial reversal learning, passive avoidance, a visual discrimination and an inclined-plane discrimination. An unusual feature of this work was the use of an inbred rat strain (Sprague-Dawley).

To produce individual differences, the inbred rats received lesions in brain areas that prior work had identified as associated with particular cognitive mechanisms. In control animals ($n = 24$) that were not lesioned, a matrix of zero-order correlations was obtained. When these control subjects were combined with all lesioned animals ($n = 96$) however, positive manifold was observed. The general factor accounted for 34% of matrix variance and the average correlation was 0.18.

Anderson (1993, 1995) developed a three-task battery using Long-Evans rats, an outbred strain, that included a novelty preference test, a type of spatial matching to sample task, and a detour test involving an inclined plane. Results indicated positive manifold, an average correlation of 0.33 ($n = 22$), and a first factor that accounted for approximately 29% of matrix variance. It should be noted that neither the Anderson nor the Thompson batteries contained control procedures to assay the impact of factors such as activity, stress, etc. on phenotypic covariance.

Plomin and colleagues reported results from a battery developed for Heterogenous Stock (HS) mice containing four spatial tasks that spanned appetitive and aversive motivational systems (Galsworthy *et al.* 2002). These tasks included spontaneous alternation, Hebb-Williams, Morris maze and a burrowing detour task that was adapted from the Thompson battery. They also developed two additional tasks using the detour arena – a context memory task and a plug puzzle. From these tasks nine measures were taken. Results revealed positive manifold, a general factor that accounted for approximately 30% of matrix variance, and an average correlation of 0.20 ($n = 40$). An additional feature of this study was an analysis of the impact of potential confounding variables in the form of anxiety, sex and outlier cases. When regressed on these variables, first component task loadings were not significantly reduced in magnitude.

Taken together these studies provide suggestive evidence that a general factor that approaches in magnitude human *g*, may indeed be present in animal problem-solving. While the studies accomplished to date may be interpreted in that direction, several cautions are in order before that conclusion can be comfortably adopted. Perhaps most importantly, these studies are quite limited with respect to the sampling of cognitive processes. All tasks are in one form or another visio/spatial in nature. With the exception of the Thompson battery, which carries with it interpretative problems owing to the use of inbred rats, the other batteries, our own included, cannot be said to provide broad-based sampling even of spatial cognition, much less processes involving other domains. While it is reasonable to suppose that a battery adequate to assess rodent cognition would have spatial performance as one of its central features, it would be unwise to rely only on a small set of spatial tasks to mark *g* without determining empirically whether other types of spatial tasks, as well as tasks taken from other domains, mark processes unrelated to the types of spatial performance thus far studied. Along these lines, Jensen & Weng (1994) argued that for the goodness of fit of a general factor, depends importantly on the number of different cognitive functions that are represented in a particular battery.

To expand the nature of the problem-solving batteries studied thus far we constructed a spatial battery that built upon our prior findings and at the same time included tasks that would measure aspects of spatial performance we had not previously assessed. The core of our battery was two spatial tasks run under water escape motivation, Hebb-Williams and place learning. In Locurto & Scanlon (1998) and in unpublished observations from our laboratory these two tasks have shown significant correlation between and within motivational conditions, suggesting that they are robust in marking some type of spatial factor. To this core, we added tasks of two sorts. The first were those that tap working memory. In the human literature it has become clear that working memory is highly related to, if not fundamental to whatever is measured by *g* (see MacKintosh 1998, Chapters 8 & 9).[1]

We developed three tasks for this battery that measure working memory. The first task was an eight-arm radial maze acquisition task. Although there have been some reports of failure using mice in radial mazes (e.g. Mizumori *et al.* 1982), in our version of this maze mice reached nearly asymptotic responding after only 10 trials (five sessions). There is a caution in identifying radial maze acquisition as a measure of working memory because it is possible for subjects to adopt search algorithms such as 'enter the next arm to the left when exiting an arm' that would lead to errorless radial maze performance without the engagement of working memory. It is apparent that rodents solve radial maze problems when these sorts of strategies are expressly prevented by manipulating aspects of the spatial environment (e.g. Olton & Samuelson 1976; see Roberts 1998, for a summary of relevant research). However, our version of radial maze acquisition used none of these manipulations, thereby allowing for the possibility of algorithmic search strategies. We therefore included in this battery a stronger

---

[1] The use of the term 'working memory' differs to some extent in the human and animal literatures. In the human memory literature short-term memory as a kind of working memory is considered a multicomponent structure, whereas in the comparable animal literature the term is often used operationally to refer to the experimental procedures (e.g. delayed matching to sample) that invoke working memory (Olton *et al.* 1979). In the Baddeley and Hitch (1994) model of human short-term memory, for example, there are separate mechanisms postulated for the short-term processing of verbal and visio-spatial information, as well as a central executive to handle scheduling and controlling functions within short-term memory. As used herein, the term refers to the common denominator between the two literatures, that is, the temporary storage and use of information for immediate problem-solving.

test of radial maze working memory that precluded successful algorithms. This task, termed the 4 × 4 test, was also run in the radial maze and followed radial maze acquisition. Our third working memory task was a visual non-matching to sample task (VNMTS) in which subjects had to use intramaze visual cues to determine correct arm choices.

Our second type of task expansion was to develop a set of detour problems that resembled the set developed by Thompson and colleagues. Our rationale for developing these problems derived from arguments made originally by Gestalt psychologists and later taken up by Tolman (1932/1967). These workers believed that the prototype of animal problem-solving was not to be found in standard protocols such as mazes. Rather, the appropriate prototype was best located in situations in which an animal was forced to restructure its perceptual field to find a previously unavailable solution to a problem. Our detour problems presented that sort of requirement to subjects.

Among these six problems three were run under water deprivation (the detour problems, radial maze, 4 × 4 test) and three were run under water escape (Hebb-Williams, place learning, VNMTS). The battery also included three control procedures: land and water activity measures, and a light-dark test.

Materials and methods

Subjects

Subjects were 60 HS mice obtained from the Institute for Behavioral Genetics, University of Colorado. There were 30 males and 30 females. These mice are an outbred strain that was originally derived from a systematic eight-way cross of inbred mouse strains (Mclearn *et al.* 1970). The mice were housed in same sex groupings of 2–4 mice in polycarbonate cages (Ancare, Bellmore, NY, USA) that measured 29.5 × 15 × 10.5 cm. Housing conditions were monitored and approved by the Holy Cross Institutional Animal Care and Use Committee. Mice were maintained on a 12-h reversed day-night cycle keyed to 6 a.m. and 6 p.m. local time. Free food (Harlan Teklad LM-485, Madison, WI, USA) and water were continuously available during the month of housing prior to the start of training and during all control procedures and water escape tasks. Mice were tested at least 2 h into their dark cycle. They were 75–90 days old at the start of training.

Apparatus and tasks

Detour problems

The four detour problems were adapted from the set developed by Robert Thompson and colleagues (Thompson *et al.* 1990). The Thompson set used three detour problems; an inclined plane, an elevated cylinder, and a ladder problem. To that set we developed and added a fourth problem, the long-ramp. The problems were run in sessions that consisted of 10 trials. On each trial a maximum of 180 seconds was allowed for subjects to find the 30 μl whole milk reinforcer. This same reinforcer was used in all water deprivation tasks. Four sessions were devoted to the inclined plane, three each to the ladder and long ramp, and two to the tube problem. The problems were run in a rectangular wooden arena that was 115 × 39 × 38 cm. One end of the arena was painted black and served as the goal area for each problem. The other end of the arena was painted off-white and served as the start area for each problem. All trials were started by lifting a floor-to-ceiling door that confined the subject to the front of the arena. In front of the goal area was a floor-to-ceiling insert with a 7.6-cm square opening cut in its center at floor level. Subjects were first trained to reach the reinforcer by entering the goal area through the square opening. Each problem divided the remaining space into different types of detours. The lighting for all detour problems consisted of one 90 watt fluorescent tube lamp positioned approximately 2.5 m above the arena and recessed into the ceiling (122 lux as measured from the center of the arena).

In the inclined plane problem, an insert 8 cm at its high end and floor level at its low end, 38 cm wide and 35 cm long, was placed with its high end away from the square opening. Subjects had to mount the plane, then descend along it into the square opening at floor level to obtain the reinforcer. A clear Plexiglas plane was used for the first session of this problem whereas the next three sessions used a metal plane. The rationale for this sequence was that in pilot testing, subjects often retreated to the darkened area beneath the metal plane if it was used first. The clear plane in the first session more readily allowed subjects to find the correct route to the reinforcer.

For this first problem and the next two detour problems (ladder and tube), any movement that resulted in the animal's body, excepting its tail, crossing and then re-crossing an imaginary line formed by the leading edge of the detour constituted an error. Errors were scored by observers who stood adjacent to this imaginary line when scoring or used a video monitor that took its feed from a video camera placed directly over the arena. The tube problem used a cardboard tube that measured 5 cm in diameter and was 6 cm long. The tube was centered 12 cm above floor level and was placed 3 cm into the goal area. It replaced the square opening as the entry to the reinforcer area. The 5-step ladder was attached to an opening cut 13 cm above the floor. This opening replaced the original square opening. An identical ladder was positioned on the other side of the opening within the goal box. The last detour problem, the long ramp, was 48 cm long and 10 cm wide. It was positioned to one side of the arena and rose to a height of 14 cm. It led to an opening cut to that side of the barrier. A smaller ramp was placed on the goal side of the barrier. The center opening of the barrier was closed for this problem. For this problem an error was recorded if the animal crossed and then re-crossed an imaginary line 27 cm beyond the leading edge of the ramp. Scores on the detour problem represented mean performance across the four problems.

Habituation to this apparatus consisted of three to four sessions, depending on how quickly subjects acclimated to the apparatus and freely consumed the reinforcer. In the first session subjects were placed in the detour box with no barrier present and were given 5 min to explore the box and consume the reinforcer. For this and the subsequent session the soda-bottle cap was filled with milk. In the next one to two sessions the floor-to-ceiling insert was present, and subjects were required to move through the square opening of the insert to receive the reinforcer that now was 30 μl.

Radial maze

An eight-arm maze constructed of black Plexiglas was used for this task. Each arm measured 31 × 18 × 31 cm. Each arm was baited at the beginning of the trial with 30 μl of milk and subjects were given 5 min to locate all reinforcers. A trial began with the subject placed in the center of the maze. Errors were defined as 15.5 cm re-entries into a visited arm or failure to consume the reinforcer in the first visit to an arm. Errors were scored by observers watching a video monitor that took its feed from a camera placed directly over the center of the radial maze. Error lines were drawn on the monitor and an error was scored whenever the mouse's body, but not its tail, fully crossed that line. The task was run for five sessions, with two trials per session. Two habituation sessions preceded acquisition. In each habituation session numerous 30 μl milk reinforcers were placed throughout the maze and subjects were given 5 min to consume them. During training each subject received two trials per session. The task was run for five sessions with indirect lighting provided by one 90 W flood lamp (105 lux).

4 × 4 test

For five sessions following the last session of radial maze acquisition, each subject was given two trials per session, with two runs within each trial, as follows: in the first run of each pair the subject was placed in the radial maze and four of the available arms were blocked by black Plexiglas inserts that were the width and height of an arm entrance. The remaining four arms were open and baited. After the subject had consumed these four reinforcers it was removed from the maze for 30 seconds and placed in a holding cage that had the same dimensions as the home cage. Bedding was provided in the cage, but no food or water was available. After 30 seconds the subject was reintroduced into the arena that now had all arms open. The four previously blocked arms now were baited and the subject was allowed to consume the remaining four reinforcers. Upon consuming the last reinforcer the subject was removed from the arena. Errors consisted of entering arms that had been previously baited on the first run of a trial, re-entries into arms visited during this second run and entries into a correct arm but failure to consume the reinforcer. The four arms that were baited on each first run were varied across trials, so that over the 10 trials of training the subject experienced 10 different correct sequences.

Hebb-Williams

The original Hebb-Williams maze (Hebb & Williams 1946) consisted of a series of 12 problems, each of which partitioned the same enclosed space so that a different route to the goal area was required. The start and goal areas remained the same on each problem. Our version of this task was run in a 51 × 51 × 16 cm arena constructed of Plexiglas with 10 cm between adjacent alleys. Our development of a water-maze version of this task (Locurto & Scanlon 1998) used five problems from Rabinovitch & Rosvold's (1951) standardization of Hebb-Williams, problems 1, 3, 4, 5 and 8. Practice problems A and D from that series were used as adaptation sessions. Errors were redefined from Rabinovitch and Rosvold's work such that each problem had four possible errors. Each problem was run for eight trials. Errors were defined as 7.6 cm entries into an incorrect turn or into a blind portion of the maze. Errors were scored by observers who stood behind the start area of the arena. Only one problem was administered to a subject each session. Subjects were given 40 seconds to find the platform. A Plexiglas platform 7.6 cm square and 11 cm high served as the escape platform. The platform was submerged just below the water line. Indirect lighting for all water escape tasks was provided by three 75 W lamps (150 lux as measured from the center of the containment pool that was used for each maze).

For this task and all water escape tasks subjects received one or two habituation sessions. In each session the platform was moved to different positions within the respective arena/maze and subjects were given 40 seconds to find the platform. The number of habituation sessions was determined by the performance of a particular group of subjects. Generally, the first water escape task experienced by subjects required two habituation sessions. Thereafter, one or two sessions were needed. In all habituation sessions, if subjects did not find the platform in that time they were placed on the platform for 10 seconds. For Hebb-Williams these habituation sessions preceded practice problem A.

Place learning

This task was conducted in a plus (four-arm) maze that was constructed of black Plexiglas. Each arm measured 46 × 31 × 15 cm. Subjects were given nine trials per session, with three blocks of three trials each session. Training consisted of 72 trials over eight sessions. Within each block of three trials subjects were started from each of three arms and had to locate the reinforcer that remained fixed during a session in the fourth arm within 40 seconds. The order of the starting positions within each block of three runs was determined randomly, with the limitation that the same sequence could not be used twice in succession within a session. This task therefore required a subject to execute a left turn, right turn and straight-alley path, randomly sequenced, within each block of three trials. For this task and all others with goal arms, an error consisted of the mouse's body, with the exception of that tail, crossing the midpoint of the goal arm. A metal platform 15 × 10 × 16.5 cm served as the escape platform.

Visual nonmatching to sample

This task was conducted in a T-maze constructed of black Plexiglas with dimensions identical to those of the plus maze used in place learning. A trial was composed of three runs: For the first two runs one of the choice arms was blocked with a piece of black Plexiglas that ran from the top of the arm to 10 cm below the water line. The block forced the subject to swim from the start arm, which remained the same on each run, to the unblocked arm. On these two sample runs the start and goal arm were the same color, either black or white. Upon finding the escape platform the subject remained there for 10 seconds. Following the second sample run the subject was given a comparison run in which both choice arms were open. On the comparison run the platform was placed in the arm with the color that did not match the color of the start arm. The color of the correct choice arm was varied left or right in a random fashion across the 10 trials that comprised a session, with the restriction that one side was correct for no more than three consecutive trials. Over the course of a session the correct non-matching color was equally often black or white. The task was run for seven sessions.

Light/dark test

This control procedure was run in a metal arena painted gun-metal gray that measured 61 cm square and 37 cm high. The arena was divided into light and dark halves. The dark half consisted of a four-sided insert made of black Plexiglas that covered one-half of the arena and the wall to the rear of the insert. The light side was illuminated with one 90-watt fluorescent lamp positioned approximately 3.2 m above the arena (155 lux). The task was run for one 10 min session. For this task we measured the proportion of time spent in the light area of the arena and the number of light-dark transitions. The task was video-taped and scored by observers for both time spent in the light portion of the arena and transitions between the light and dark areas.

Activity controls

Both activity control procedures were run in a square arena constructed of Plexiglas and painted flat ivory that measured 61 cm on each side and was 46 cm high. For the land activity task subjects were given one 10 min session in the same lighting as was provided for the detour task. For the water-based activity measure the arena was filled to a depth of 16 cm and subjects were given 5 min in the arena under the same lighting provided for all water escape tasks. Each control session was separated from the next by 24 h. The water activity tests was videotaped. Observers scored performance for total distance by dividing the arena into nine equal squares. The number of squares crossed was used as the dependent measure. The land activity test was scored by Ethovision (Noldus, Wageningen, the Netherlands). For the land activity procedure total distance as well as thigmotaxis was recorded. Thigmotaxis was defined as movement within a 4.5-cm zone measured from the borders of the apparatus. These control procedures yielded five measures, two each from the land activity and the light/dark test and one from the water activity test.

General procedures

Subjects were run in squads of four such that each subject was given the first trial before the second trial had begun. This procedure had the effect of insuring that the average intertrial interval ranged on average between 60 and 120 seconds for each problem. Between trials subjects were placed in holding cages. For water escape tasks these cages consisted of standard-sized mouse cages with paper bedding on the floor. For water deprivation tasks bedding was provided on the floor. No food or water was available in these cages.

The water escape tasks were run as has been described in Locurto & Scanlon (1998). Briefly, water temperature was maintained at 26 ± 1 °C. In each task subjects were started with their head facing the back wall of the arena/maze. The water was whisked between trials to remove odor cues, and the water was partially replaced as needed to maintain water temperature.

Water deprivation was achieved by allowing each mouse to drink in its home cage for a period of 30 min following experimental sessions and on days when experiments were not run. Over a period of five days the amount of time allowed for drinking was reduced from 2 h to 30 min. This schedule maintained mice at approximately 85–90% of their *ad libitum* weights and provided for approximately a 21-h deprivation condition for each session. If a subject was not maintaining this weight level it was given 15–30 additional min of water access in a cage separate from its home cage with food available. For all water deprivation tasks the arenas were cleaned between trials with a 50/50 water/ethanol solution.

Subjects were divided into three subgroups. One subgroup experienced the water deprivation tasks first. The second group experienced the water escape tasks first. The third subgroup experienced two olfactory tasks run under food deprivation first and then the water deprivation tasks followed by water escape. The olfactory tasks were in the process of development within this one subgroup and the results for these tasks are not reported herein. Within a motivational condition, tasks were run in a constant order. For water deprivation the order was the detour problems followed by the radial arm maze acquisition task and the 4 × 4 test. For water escape the sequence for all subjects was Hebb-Williams, place learning and VNMTS. Tasks were separated by no more than one week within a motivational condition. Three to four weeks separated tasks between motivational conditions.

Dependent measures

Research assistants recorded three dependent measures: errorless trials per session, mean errors per trial and mean latencies per session. Rater reliability was greater than 0.85 for all three measures. In Locurto & Scanlon (1998), errorless trials and latencies showed only moderate covariance, suggesting that each measure assayed a somewhat different aspect of performance. Errors were added to this study because subsequent analyses of those earlier data showed only moderate correlations between errors and these other two measures. Other research programs that developed batteries of cognitive tasks have at times used predominantly latencies (e.g. Galsworthy *et al.* 2002), a combination of errors and latencies (Anderson 1995) or have reported each measure separately (Locurto & Scanlon 1998). It is not possible based on prior evidence to decide whether one dependent measure or another serves as the best index of cognitive performance. Our lab has favored tasks that yield both error and latency measures. For that reason we have not included Morris maze in the present battery, despite the fact that it may currently be the most popular task to measure spatial behavior, because as typically run it yields only latencies.

For the present study we derived an aggregated measure composed of the three dependent measures. The use of this aggregated measure offered two advantages. First, this measure provided consistently higher reliability for each task than did the average of the univariate dependent measures for that task. Second, aggregated scores weighed all facets of performance equally, obviating the need to make decisions as to which univariate measures to rely on for analysis and interpretation. Aggregated scores were obtained by standardizing the raw scores for each measure and then summing these standardized scores. Before summation, errorless trial scores were multiplied by -1 to convert them to the same scale-direction as latencies and errors, meaning that higher scores represented poorer performance. Following presentation of the descriptive statistics for each measure, the remaining results, including all correlation and data reduction analyses, are presented using this aggregated measure. The Appendix provides these same analyses using the three univariate measures separately.

Data analysis

All analyses were conducted using SPSS v10.7 (http://www.spss.com). Data reduction was accomplished via principal components. Principal components was chosen because the central question in this study concerned the appropriate number of components/factors to be extracted from our data set. Principal components is designed to answer questions of data reduction, that is, into how many components should a large data set be reduced (Dunteman 1989)? In cases where more than one component with an eigenvalue exceeding 1.0 was derived from the unrotated solution, both orthogonal (Varimax) and oblique (Direct Oblimin) rotations were performed. These rotations did not produce different solutions for any analysis. In all cases we report the orthogonal rotation. We also report component correlations from the oblique rotation to indicate commonality between components. The unrotated solutions for each data reduction analysis involving the problem-solving tasks are reported within the text. It is commonplace in studies of the general factor in human intelligence to use unrotated solutions because rotations of any sort have the property of redistributing variance from the first principal component to other components. The unrotated solution therefore often provides a preferred estimate of the amount of variance accruing to the first principal component.

Results

Descriptive statistics

Table 1 presents the means and standard deviations for errorless trials (ETS), errors and latencies for each problem-solving task. To give a sense of acquisition, the first and last portions of exposure to each task are presented and paired t-statistics were used to gauge the magnitude of the changes across training. For the detour problems performance was summed across problems, meaning that the term ‘first’ in the table corresponds to performance on the first session of each problem summed across different problems, while the term ‘last’ refers to the summed performance across problems for the last training sessions. By this measure subjects averaged 4.34 errorless trials on their first session of training across the detour problems, and 6.71 errorless trials on the last sessions (out of 10 trials). For Hebb-Williams, data were summed across trials, given that each problem is presented for only one session. As a result, the terms ‘first’ and ‘last’ refer to performance on the first and last trial of each problem summed across problems. For the remaining tasks, problem solution remained the same throughout training, allowing direct comparisons between first and last sessions on the same problem. As can be seen in the table, all tasks showed increases in the frequency of errorless trials and decreases in errors and latencies. In all but two of the 21 comparisons, VNMTS for errors and place learning for latency, these changes reached statistical significance.

The analysis of sex differences for the 18 possible measure × problem-solving task comparisons (six tasks, three dependent measures) revealed four significant sex differences: for 4 × 4 latencies and for radial arm maze, 4 × 4 and VNMTS errors ($P < 0.05$ for each comparison; lower latencies and fewer errors for females in each case). As indicated in the data reduction analyses given below, when sex was entered as a dummy variable in the principal components analysis it produced no significant loadings on any components. Regression analyses also indicated no

Table 1: Descriptive statistics for problem solving tasks

		First	Last			First	Last
ETS							
Detour[a]	mean	4.34	6.71**	HW[b]	mean	0.11[c]	0.76**
	SD	2.73	2.59		SD	0.31	0.43
RAM	mean	4.78	6.37**	Place	mean	4.45	5.93**
	SD	1.94	1.82		SD	1.97	2.22
4 × 4	mean	0.85	1.70**	Visual NMTS	mean	4.93	6.04**
	SD	1.00	0.81		SD	1.46	1.72
Errors							
Detour[a]	mean	1.46	0.64**	HW[b]	mean	4.14	0.42**
	SD	1.15	0.99		SD	4.61	0.86
RAM	mean	4.83	1.85**	Place	mean	1.51	1.03**
	SD	3.88	2.18		SD	1.34	1.97
4 × 4	mean	4.83	3.27**	Visual NMTS	mean	1.31	1.11
	SD	2.95	1.93		SD	1.87	2.14
Latency							
Detour[a]	mean	42.70	23.74**	HW[b]	mean	26.59	11.50**
	SD	29.00	27.20		SD	12.15	9.07
RAM	mean	95.46	57.58**	Place	mean	13.19	12.72
	SD	38.00	30.44		SD	6.32	8.86
4 × 4	mean	64.29	46.96**	Visual NMTS	mean	11.56	9.03**
	SD	41.56	17.95		SD	5.53	4.25

**$P < 0.01$; [a]Data are collapsed across problems; [b]Data are collapsed across trials; [c]As data are summed across trials, the ETS measure is the proportion of trials without an error.

significant effect of sex on individual task aggregated scores with the single exception of the 4 × 4 task (t(53) = 2.15, $P = 0.025$; R = 0.30). For that task the residuals resulting from regressing aggregated scores on sex were used in a principal components analysis separate from the primary analysis to determine the effects of this sex difference on component structure. That analysis is reported below.

Table 2 provides correlations between the univariate dependent measures and task reliabilities using Chronbach's alpha (α) for each dependent measure and for the aggregated measure. The correlations between errorless trials and both latencies and errors were uniformly negative, averaging −0.48 (latency) and −0.62 (errors). The average correlation between errors and latencies was 0.44. Task reliabilities averaged 0.66 for errorless trials, 0.88 for errors and 0.85 for latencies. In prior work published from our laboratory, latency reliabilities have consistently exceeded errorless trial reliabilities. The magnitude of error reliabilities suggests that this measure parallels or exceeds latencies in terms of consistency. As can be seen in the right-hand column in Table 2, the reliabilities for the aggregated scores were uniformly high for each task, and averaged 0.84. These aggregated scores were used in all correlational and data reduction analyses.

Table 3 presents the analysis of order effects for each dependent measure. Order effects represent differences in mean performance between the three groups that were exposed to the tasks in counterbalanced orders. These differences were tested with analyses of variance followed by Tukey's HSD test. It can be seen that in six of the possible 21 task × dependent measure comparisons, order effects reached statistical significance. The effects of order were not consistent across dependent measures or tasks, however. In the detour task, experienced subjects recorded significantly higher errorless trials, but not fewer errors, than did experimentally naïve subjects. Also for this measure, subjects receiving this task first, recorded higher latencies than did subjects who experienced this task second. For the radial maze, experienced subjects produced lower latencies and fewer errors than naïve subjects, but order did not affect errorless trials. For VNMTS subjects, experiencing these tasks first produced higher levels of errorless trials and fewer errors than did experienced subjects. It was also the case that order had no effect on any measure for three tasks, the 4 × 4 working memory test, Hebb-Williams and place learning.

These different patterns suggest that while order effects may occasionally be obtained in battery work of this sort, the effects of order interact with the task, dependent measure and history under consideration, rather than constituting a uniform progressive error that produces the same effect across all tasks and all dependent measures. The effects of order were directly tested in one of the data reduction analyses reported below by standardizing scores within orders and within tasks to eliminate central tendency differ-

Table 2: Correlations (r) between problem solving measures, and measure reliabilities (α)

	r			α_r	$\alpha_{agg.}$
	1	2	3		
Detour					
1. ETS	1.00			0.90	
2. Errors	-0.75^{**}	1.00		0.88	0.93
3. Latency	-0.86^{**}	0.78^{**}	1.00	0.92	
Radial Arm Maze					
1. ETS	1.00			0.66	
2. Errors	-0.69^{**}	1.00		0.87	0.82
3. Latency	-0.51^{**}	0.71^{**}	1.00	0.72	
4×4					
1. ETS	1.00			0.68	
2. Errors	-0.51^{**}	1.00		0.78	0.81
3. Latency	-0.13	0.53^{**}	1.00	0.80	
Hebb-Williams					
1. ETS	1.00			0.56	
2. Errors	-0.59^{**}	1.00		0.89	0.80
3. Latency	-0.51^{**}	0.24^{*}	1.00	0.91	
Place					
1. ETS	1.00			0.77	
2. Errors	-0.58^{**}	1.00		0.91	0.84
3. Latency	-0.76^{**}	0.40^{**}	1.00	0.85	
Visual NMTS					
1. ETS	1.00			0.38	
2. Errors	-0.31^{*}	1.00		0.92	0.85
3. Latency	-0.23	0.09	1.00	0.87	

$^{*}P<0.05$; $^{**}P<0.01$.

Table 3: P values for order effects

Task	ETS	Errors	Latency
Detour	^{*a}	ns	^{*b}
RAM	ns	^{*c}	^{*d}
4×4	ns	ns	ns
HW	ns	ns	ns
Place	ns	ns	ns
Visual NMTS	^{*e}	^{*f}	ns

ns = non-significant; $^{a}=1<2$ & 3; $^{b}=1>2$; $^{c}=1>2$ & 3; $^{d}=1>2$ & 3; $^{e}=1>3$; $^{f}=1<2$ & 3. Numbers refer to ordinal position of task. $^{*}P<0.05$.

ences. These scores were then submitted to principal components analysis.

Correlational analyses and data reduction

Prior to conducting correlational analyses for the problem-solving tasks and control procedures, the five control measures were submitted to a principal components analysis using orthogonal rotation. This analysis attempted to reduce these five measures to a smaller subset of components that might be more easily interpreted. That analysis revealed two distinct components ($r=-0.07$ from the oblique rotation) that together accounted for nearly 67% of matrix variance. The first (rotated) component, accounting for 37.49% of matrix variance, was marked principally by light-dark transitions (a loading of 0.78) and the land distance measure (0.76).[1]

We named this factor 'Activity.' The second rotated component, accounting for 28.74% of matrix variance, was marked principally by the water activity task (0.89), and to a lesser extent by the Thigmotaxis measure from the land activity task (0.60). We named this component 'Stress.' These summary components were then substituted for the five control measures in correlational and data reduction analyses. Scores on these two components were generated by weighing subjects' standardized scores on each component by the loading of each control measure on that component, and then summing these weighted scores. The correlation between these two sets of component scores was 0.02, indicating that the independence of these two components that was evident in the principal components analysis was preserved in the constructed component scores.

Two aspects of this control procedure analysis should be noted. First, the proportion of time spent in the light area of the light-dark arena, a measure often used as a measure of anxiety (see Hascoët *et al.* 2001 for a review) did not correlate significantly with any other control measure, nor did it load on either of the first two factors. This task was therefore excluded from the final principal components analysis that produced the two control components. Second, it might be expected that land and water activity measures would correlate given that they were run in the same apparatus and putatively are measuring similar constructs involving overall movement/activity. However, the correlation between these two measures was 0.06, and they marked independent components.

Table 4 presents the correlation analyses for aggregated scores across all problem-solving tasks and the summary control components. This same analysis for each dependent measure considered separately along with the original five control measures is given in the Appendix, Table A1. It can be seen in Table 4 that the majority of problem-solving correlations were positive. Six of these correlations were, however, of small magnitude, arbitrarily defined as $0.00\leq r\leq 0.15$, and four of the correlations, involving the VNMTS and the 4×4 tasks, were negative. Neither of the control components bore a positive relationship with any of the problem-solving tasks.

---

[1] A common criterion for interpreting orthogonal component/factor loadings as significant is 0.71. This criterion derives from the idea of treating these loadings much like correlation coefficients. Consequently, the square of 0.71, 0.50 indicates that a particular variable shares 50% of its variance with the factor in question. Similarly, a common cutoff to define marking variables as 'poor' fits with a factor is 0.30 or less, given that a variable with this loading would share less than 10% of its variance with the factor (Comrey & Lee 1992, p. 242–244).

Table 4: Correlation matrix for aggregated scores and control components

	1	2	3	4	5	6	7	8
1. Detour	1.00							
2. RAM	0.26*	1.00						
3. 4 × 4	0.21	0.14	1.00					
4. HW	0.12	0.04	−0.11	1.00				
5. Place	0.11	0.11	−0.01	0.46**	1.00			
6. VNMTS	−0.13	−0.04	0.35**	0.13	0.18	1.00		
7. Activity	−0.04	−0.04	−0.04	−0.17	0.00	−0.09	1.00	
8. Stress	−0.03	−0.11	0.16	0.00	−0.03	−0.11	0.02	1.00

*$P < 0.05$; **$P < 0.01$.

Table 5 presents the results of applying principal components analysis to this correlation matrix and Table 6 gives the component correlations derived from the oblique rotation. The comparable analyses of the univariate data can be found in the Appendix, Tables A2 and A3. Orthogonal rotation was used because the non-rotated solutions for both aggregated and univariate scores each reduced to multiple components without an over-arching first component. For the data in Table 5, the variance explained by the first component in the unrotated solution was 22.19%, compared to 19.43% for the orthogonal solution. This same pattern obtained for both orthogonal and oblique rotations. Additionally, the orthogonal and oblique rotations produced nearly identical component loading patterns for each dependent measure.

The results in Tables 5 and 6 taken together indicate that the matrix of six problem-solving tasks and two control components reduced to four independent components whose eigenvalues exceeded 1.0. The first component was marked by Hebb-Williams (0.87) and place learning (0.80). The second component consisted of the detour problems (0.80) and the radial maze acquisition procedure (0.74). The third component was marked by two of the working memory tasks, VNMTS (0.83) and the 4 × 4 task (0.79). The fourth component was marked almost exclusively by the Stress component (0.95). The Activity component did not appear to mark a factor, although the loading pattern in the first three components – highly positive loadings for some problem-solving tasks and a negative loading for Activity within each component – might be interpreted as representing the distinction between problem-solving performance and Activity.

Table 5: Rotated component (C) loadings, eigenvalues and percent variance for aggregated scores and controls

	C1	C2	C3	C4
Task				
Detour	0.12	0.80	−0.02	0.13
RAM	0.03	0.74	0.06	−0.23
4 × 4	−0.15	0.33	0.79	0.29
HW	0.87	0.02	−0.06	0.04
Place	0.80	0.09	0.06	−0.02
VNMTS	0.23	−0.25	0.83	−0.21
Activity	−0.25	−0.04	−0.19	−0.04
Stress	0.00	−0.08	0.00	0.95
Eigenvalue	1.55	1.37	1.35	1.10
% Variance	19.43	17.12	16.98	13.70

The effects of sex differences on this analysis were tested in two ways. In the first, sex was entered as a standardized binary dummy variable in the principal components analysis. The addition of sex produced a five factor solution with explained variance ranging from 17.93 to 11.78% across components. That analysis indicated that sex produced no significant positive loadings on any component. The loadings for sex ranged from −0.64 on C1 that was marked in this analysis by the 4 × 4 task and VNMTS, to 0.32 on C2 that was marked by Hebb-Williams and place learning. In the second analysis the 4 × 4 task, the only task that evidenced significant regression for sex, was regressed on sex and the residual aggregated scores were used in a principal components analysis in place of the original 4 × 4 scores. That analysis revealed the identical component structure that had resulted from using the original scores. Four components emerged with explained variance estimates ranging from 19.52 to 13.49%. The third component was again marked by the 4 × 4 task and VNMTS.

To test the effects of order on these analyses, aggregated scores were first standardized within the three subgroups that experienced the tasks in counterbalanced orders. That is, scores were standardized within each subgroup only against other scores within that subgroup. The scores from all subgroups combined were then submitted to principal

Table 6: Component correlations from oblique rotation of aggregated scores

	1	2	3	4
1. C1	1.00			
2. C2	0.05	1.00		
3. C3	0.15	0.07	1.00	
4. C4	−0.11	0.03	−0.03	1.00

components analysis. The component structure resulting from that analysis was highly similar to the analysis presented in Tables 5 and 6. Four components were extracted with eigenvalues exceeding 1.0. The variance accounted for by the rotated components ranged from 18.68 to 14.82%. The unrotated first principal component accounted for 20.48% of matrix variance. The first component was again marked strongly by Hebb-Williams (0.80), and there was a separate component marked almost exclusively by Stress (C3, 0.77). Component correlations from the oblique analysis indicated that the four components were independent (*r* range: −0.08 to 0.08). There were some task-loading differences between this analysis and the one presented in Table 5. Place learning loaded more modestly on the first component in this analysis than in the original aggregated score analysis (0.22 vs. 80), and the detour task loaded more highly on the first component in this analysis than in the original analysis (0.65 vs. 12).

Alternative structures

The analyses presented thus far for univariate and aggregated scores were all based on averaging each subject's performance across all trials of a given task to produce an overall summary score. There are at least two alternative approaches to analysis of this data set. First, it might be argued that extended training on a task introduces differences in attentional and motivational factors that may obscure the search for general cognitive ability. Our strategy has been to run subjects until performance surpasses chance and meets common standards of acquisition. If extended training has these aforementioned effects it might be efficacious to terminate training on each task at much earlier stages. While fully implementing this strategy would require a design in which each task was run for only a short period of time, with the present data set we can look at the structure of performance after a brief exposure to each procedure. Second, taking just the opposite tack, it might be argued that performance early in training is influenced initially by differences in stress, activity, etc., factors that may dissipate with extended training. It might be argued that the proper measure of cognitive ability comes from focusing on terminal levels of responding following extended training, rather than summing together all aspects of training.

Both of these analyses were performed using aggregated scores. For the early training analysis, performance during the first session was used. For late training scores, performance during the last session was used. The results of orthogonal rotation of the early and late correlation matrices is presented in Table 7. For the early training, four components emerged and the percentage variance accounted for ranged from 22.94 to 13.60%. The unrotated first principal component accounted for 23.32% of matrix variance. Correlations from the oblique rotation were all negative and ranged from −0.01 to −0.07. For the analysis of late training scores, once again four components emerged, and variance accounted for by these components ranged from 20.63 to 13.37%. The unrotated first principal component for the late analysis accounted for 21.74% of matrix variance. Component correlations ranged from −0.05 to 0.11.

Despite the similarities in component structure evident in these two analyses, there were differences between them in terms of task loadings that suggest trends in problem-solving structure over the course of extended training. The first component, for example, was marked early in training by Hebb-Williams, place learning and the radial maze. Late in training the first component was marked more strongly by Hebb-Williams and the detour tasks. Place learning was associated with the VNMTS task late in training and marked the third component, while the radial maze task along with the 4 × 4 task marked the second component late in training. Additionally, while the two control components did not affect problem-solving when assessed via overall aggregated scores, Stress appears to have covaried with the two working memory tasks, 4 × 4 and VNMTS (C2), early in training. Late in training Stress was more associated with the radial maze and 4 × 4 test (C2). Activity bore a negative relationship to

Table 7: Rotated Component Loadings, Eigenvalues, and Percent Variance Accounted for Using Early and Late Aggregated Scores

	Early				Late			
	C1	C2	C3	C4	C1	C2	C3	C4
Task								
Detour	−0.05	0.04	0.89	0.01	0.90	0.07	−0.09	0.11
RAM	0.75	0.01	−0.05	−0.37	0.27	0.67	−0.05	0.21
4 × 4	−0.17	0.77	0.36	−0.09	−0.09	0.83	0.05	0.01
HW	0.72	0.12	−0.21	0.13	0.79	−0.11	0.35	−0.27
Place	0.75	−0.18	0.21	0.12	0.11	0.24	0.74	−0.04
Visual NMTS	0.35	0.76	0.02	−0.03	−0.03	−0.11	0.83	0.21
Activity	−0.02	0.05	0.00	−0.95	−0.04	0.07	0.14	0.92
Stress	−0.22	0.70	−0.36	0.02	−0.35	0.54	0.24	−0.23
Eigenvalue	1.83	1.71	1.15	1.08	1.65	1.52	1.46	1.07
% Variance	22.94	21.40	14.30	13.60	20.63	19.09	18.21	13.38

Hebb-Williams and place learning early in training (C4), and constituted its own independent component (C4) late in training.

Discussion

The primary goal of this work was to explore the structure of individual differences in mouse problem-solving, with the intent of observing whether or not a general cognitive ability might be detected, one that might parallel in its structure the ubiquitous general factor found in human intelligence. Examination of the principal components analyses in this study do not lend support to the assertion that these data are best characterized by the presence of a general cognitive ability. The first extracted principal component, which would represent this general factor if positive manifold were present, was uniformly of relatively modest magnitude, and was part of a set of components with nearly equivalent explained variance characteristics. This central finding was not dependent on the particular analyses performed, nor was it confined to particular measures or aggregations of measures. It was obtained when confined to univariate scores (Appendix), to aggregated scores across dependent measures, to aggregated scores derived from early or late training, as well as to aggregated scores standardized within training orders.

Modularity

These data might encourage a more modular conception of the structure of individual differences in mouse problem-solving. With respect to the distinctions between general and modular approaches, it should be noted that these two approaches need not be irreconcilable. This same distinction emerged in the human intelligence literature when Thurstone (1938) elaborated his theory that human intelligence consisted of seven primary mental abilities without *g*. It was eventually shown that his theory of multiple factors was commensurate with a general factor approach. The framework for that resolution was a hierarchical model in which Thurstone's separate abilities constituted lower-order factors in a structure that had *g* as its apex. The fact that *g* could be extracted along with lower-order factors rested on the finding that these lower-order factors themselves correlated, suggesting a source of common variance not tied uniquely to these factors (see Locurto 1997 for a discussion of this controversy).

The problem with applying this approach to our data is that the usual foundation upon which a hierarchical structure is built is a matrix of positive correlations – across tasks and components. Thurstone (1934) himself readily acknowledged that all mental tests, as they were called at the time, were positively correlated despite his insistence that the resultant matrix was best described as a set of separate primary abilities. In his initial exposition of his approach he gave a battery of 56 mental tests to a sample of 240 college students (Thurstone 1938). This battery yielded 1596 correlations with a mean of 0.35. Of those correlations, only 41, or 2.5%, fell between −0.25 and 0.00. The remaining 1555 correlations ranged between 0.01 and 0.85. Two-thirds of these correlations fell above 0.25. Additionally, in his unrotated factor matrix all 56 tests loaded positively on the first extracted factor.

Compare that pattern of correlations with the pattern observed in this study, where for aggregated scores (Table 5), 10 of the 15 problem-solving correlations were either negative or of small magnitude ($r < 0.15$), and the average correlation was 0.12. Admittedly, the power of this study is far lower than Thurstone's original study, or for that matter, most human psychometric studies of intelligence, both in terms of sample size and number of variables (tasks). Despite these limitations the differences between the two sets of data are rather striking, and they suggest at least two possibilities. One possibility is that this particular battery of animal problem-solving tasks is highly unrepresentative of animal cognition generally, and therefore it obscures what would otherwise be a matrix of positive correlations that yields a strong first principal component. The second possibility is that there may be fundamental differences in the structure of human vs. animal cognition.

The first possibility is highly unlikely. While no one battery can claim to represent all aspects of animal problem solving, this battery does include standard measures of spatial learning (Hebb-Williams, place learning, radial maze) and working memory (4 × 4 test and VNMTS). Additionally, the detour problems have a well-established rationale for their inclusion in this battery, as described earlier.

The second possibility, that animal and human intelligence differ fundamentally, is intriguing and comports with modern analyses of the uniqueness of human intelligence, especially the distinctive dominance of language (Macphail 1982; Deacon 1997). Yet it is not prudent to draw any strong conclusions from our data about that possibility. The confirmation of a modular structure in animal problem-solving would involve the study of a rather broad array of task types. The finding of modularity in one study, no matter how well-founded the rationale for task selection, does not preclude the possibility that in examining another set of tasks, especially those that measure other dimensions of animal performance, a broader set of commonalities might reveal themselves (see Jensen *et al.* 1978, for similar findings using a different problem battery).

Psychometric considerations

The number of components extracted in this study is large given the number of tasks used. Stated differently, principal components analysis did not produce much data reduction: nine tasks (eight variables) were reduced to four components. There may be a number of reasons for this failure to reach greater simplicity in data reduction. The first and foremost reason typically would be unreliability of the data set. In this study, however, reliability would appear to be a strength, not a weakness, especially with respect to the aggregated scores that were used for all primary analyses. The average reliability of these aggregated scores was 0.84. In comparison, human intelligence tests average approximately 0.90, and standardized achievement tests average 0.85. Classroom multiple-choice tests and rating scales average between 0.70 and 0.75 (Murphy & Davidshofer 1988). Moreover, Nunnally (1967) has noted that attempts to increase reliabilities above 0.80 are wasteful in that little if any additional random error will be removed in the process.

The presence of a relatively large number of components also makes interpretation difficult. While it is clear that the first extracted component does not represent a single overarching general factor, it is not a simple matter to interpret so many independent components given the presence of relatively few tasks. The suggestion from these data may be that while there is some coherence among tasks as marked by components, much of the variance in the correlation matrix is task-specific. That variance is almost certainly not due largely to random error given the magnitude of the task reliabilities for aggregated scores. What is left is variance specific to the tasks themselves.

The idea that much of the variance is tied to specific tasks and not to common components can be seen in several ways. One is to examine how much of the total matrix variance is captured by the extracted components. In Table 5 the four extracted components account cumulatively for approximately 67% of matrix variance. Fully one-third of matrix variance is not tied to any of the extracted components, despite the relatively large ratio of components to tasks. A second way is to examine the magnitude of the correlations presented in Table 4. As noted earlier, the average matrix correlation is 0.12. This average correlation suggests that while there is some covariation among variables, much of the variance attached to these variables is not expressed in correlations, that is, variance attached to one task is not well predicted from knowledge of other tasks.

It should be emphasized that calling a significant portion of the unexplained variance in this study task-specific does not necessarily mean that it is variance directly associated with cognitive processes. This variance may in large measure be arena-specific, perhaps in the nature of particular levels of activity or stress associated with particular arenas. There may also be subject–arena interactions in that some arenas evoke more activity or stress for some subjects than others. In either case, that variance would not be shared across tasks and would not be part of the variance observed in correlations.

Component structures

The discussion of variance specificity should not preclude recognition that significant components did emerge in this study. As is true in any data reduction analysis, the precise nature of extracted components depends on the types of tasks and measures included in a battery, as well as on the type of analysis undertaken.

These caveats aside, there are several conclusions worth drawing about these results apart from the absence of a general factor. In this study, as in Locurto & Scanlon (1998), Hebb-Williams and place learning marked a common component. This component has emerged despite the presence of a varying number and type of other tasks in these two batteries. It also emerged in both the aggregated scores and univariate analyses in the present study. Clearly, these two tasks appear to be marking some sort of spatial factor.

These results also suggest that what is termed spatial behavior may not be unitary in nature. In this study the detour and radial maze tasks as well as the 4 × 4 test appeared to mark components that were independent of the component marked by Hebb-Williams and place learning. This independence of different spatial tasks has important ramifications for our understanding of spatial behavior. If different studies use different tasks to study spatial behavior, and if these tasks themselves mark independent aspects of spatial behavior, then our ability to make sense of the literature and to form valid conclusions about what we are calling spatial behavior is severely limited.

As a caveat, it might be argued that the different spatial tasks coalesced around common motivation. Hebb-Williams and place learning were both run under water escape, whereas the detour and radial maze tasks were run under water deprivation. Common motivation cannot be ruled out in explaining the shared variance between some of these spatial tasks. It is evident, however, that common motivation does not explain several features of these data. The component marked by two of the working memory tasks was composed of one water escape (VNMTS) and one water deprivation task (4 × 4). Also the early and late analyses suggest that during different points in training, tasks coalesced for reasons that are not simply motivational. Early in training the first component was marked by the radial maze (water deprivation), Hebb-Williams (water escape) and place learning (water escape). Late in training, this same component was marked by the detour task (water deprivation) and Hebb-Williams.

Moreover, the presence of motivational and apparatus commonality did not produce common factor loadings for the radial maze and the 4 × 4 test. While these two tasks did evidence positive correlation, they marked different components. This difference suggests that at least for some subjects radial maze acquisition may have encouraged the use of non-working memory strategies such as search algorithms, whereas those same strategies were prevented in the 4 × 4 test and would not have been useful in VNMTS.

The common loadings of these two working memory tasks despite motivational differences is of interest because there has been considerable speculation in the human intelligence literature, well-summarized by MacKintosh (1998, see Chapters 8 and 9), that working memory may be an important component of general intelligence in humans, if not the fundamental process that underlies general intelligence. The fact that these two procedures loaded on a component that was independent of the two spatial components suggests that working memory may constitute a separable process in animal problem-solving. Much work needs to be done before the contributions of working memory are fully understood. It is not evident, for example, whether working memory should be regarded as a unitary process in animal problem-solving. In human intelligence, several forms of working memory have been identified (e.g. visual, verbal, etc.; Baddeley 1986; MacKintosh 1998). Whether an analogous partitioning is also true of animal working memory awaits appropriate study. Individual difference paradigms are ideal for this work.

It should be added that the components identified in this study, as well as in our prior work, do not appear to be dependent to any significant degree upon activity or stress, at least given how these constructs were operationalized in this study (see Locurto & Scanlon 1998; Galsworthy *et al.* 2002, for a similar finding). While more precise delineation of the structure of mouse problem-solving awaits further work, it appears that this work is not hindered by these potential artifacts.

'Constructing' g

The thrust of the analyses in this study has pointed to the presence of multiple components, each accounting for approximately the same order of magnitude of variance. While this pattern is not congenial to the presence of an over-arching general factor, in one important respect these analyses do not parallel the construction of *g* in the human intelligence literature. The development of a standardized human intelligence battery begins with a large set of items that are systematically reduced via sampling and data reduction procedures to a smaller set. This process may continue through several rounds of item sampling with different subject samples until a final battery results. That battery is composed of those items that have most strongly correlated throughout the development of the battery. The coherence across items that is seen in human batteries, is therefore the result of a deliberate, inductive strategy of parsing a large item set into a much smaller set that *per force* shows positive manifold. Items that do not load on the first component are removed.

That same strategy can be followed in the present data set by selecting a subset of tasks that showed common variance. Using aggregated scores and adopting the criterion of a 3-task battery as the minimum acceptable battery, the best subset of tasks consisted of the VNMTS task, Hebb-Williams and place learning. When these three tasks alone constituted the test battery, only one component emerged with an eigenvalue greater than 1.0. That component accounted for 51% of matrix variance. All matrix correlations were positive and the mean correlation was 0.26. An even stronger battery resulted from using the univariate measures of VNMTS errorless trials, and the error measures from radial maze, Hebb-Williams, and place learning. That four-task battery accounted for 55% of matrix variance. Again, all matrix correlations were positive, and averaged 0.39.

These average correlations and levels of variance accounted for, not including the control components, are appreciably more powerful than the batteries reported by either Anderson (1993), Galsworthy *et al.* (2002), or our earlier battery composed of water escape tasks (Locurto & Scanlon 1998). The variance explained by the first principal component in the three-task Anderson battery, which included no control procedures, averaged approximately 29%, and the average correlation was 0.33 ($n = 22$). The comparable figure for the Galsworthy *et al.* battery, which contained four independent tasks from which nine measures were taken, was approximately 30% and the average matrix correlation is this study was 0.20 ($n = 40$). For the seven-task battery devised by Thompson and colleagues, which also did not contain control procedures, analysis of lesioned and unlesioned rats combined revealed a first principal component that accounted for 34% of matrix variance, and the average correlation was 0.18 ($n = 120$; Crinella & Yu 1995).

The magnitude of explained variance in the first component of our constructed *g* batteries is also comparable to the findings in standardized human intelligence tests. For example, Miller & Vernon (1992; Table 3, p. 18) reported the results of administering the Multidimensional Aptitude Battery (MAB) to 170 college students. The MAB consists of 10 subtests that measure both verbal (e.g. vocabulary, comprehension) and non-verbal (e.g. picture arrangement, object assembly) abilities. This sample yielded an average subtest intercorrelation of 0.29 (range = 0.06 to 0.62) and a first common factor (*g*) that accounted for 37.0% of matrix variance. This result falls within the range reported by Jensen (1987) in an analysis of *g* magnitude involving 20 studies. *g* variance ranged from 33% to 61% in these studies. The variance values for our constructed batteries, it should be added, were taken from the unrotated solutions. As noted earlier, rotation of component or factor matrices typically has the effect of reallocating some of the variance that is initially allocated to the first principal component to other components or factors (see Galsworthy *et al.* 2002; Locurto & Scanlon 1998, for use of this technique in an animal battery.).

The importance of our constructed batteries lies not in their favorable comparison with other animal batteries, or with human intelligence batteries, but in their comparison with the larger battery from which they were derived. If these constructed batteries were all that was reported, they would suggest the presence of a substantial general factor, one that clearly rivals human *g* in its psychometric properties. When seen in the context of the larger variable set from which they were drawn it is evident that the *g* derived from these limited batteries fails to represent fairly the complexity of the larger and more modular component structure. This same point might be kept in mind when evaluating the results from any battery that is reported to yield positive manifold and a general factor in animals, but that is not demonstrated to be representative of a broader sampling of task types and motivational systems.

References

Anderson, B. (1993) Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: Intelligence? *Neurosci Lett* **153**, 98–112.

Anderson, B. (1995) Dendrites and cognition: a negative pilot study in the rat. *Intelligence* **20**, 291–308.

Baddeley, A.D. (1986) *Working Memory*. Oxford Science Publications, Oxford.

Baddeley, A.D. & Hitch, G.J. (1994) Developments in the concept of working memory. *Neuropsychology* **8**, 485–493.

Brody, N. (1992) *Intelligence*, 2nd edn. Academic Press, New York.

Comrey, A.L. & Lee, H.B. (1992) *A First Course in Factor Analysis*, 2nd edn. Erlbaum, Hillsdale, NJ.

Crinella, F.M. & Yu, J. (1995) Brain mechanisms in problem solving and intelligence. A replication and extension. *Intelligence* **21**, 225–246.

Deacon, T.W. (1997) *The Symbolic Species: the Co-Evolution of Language and the Brain.* Norton, New York.

Dunteman, G.H. (1989) *Principal Components Analysis*. Sage Publications, London.

Eysenck, H.J. (1987) The several meanings of intelligence. *Behav Brain Sci* **10**, 663.

Galsworthy, M.J., Paya-Cano, J.L., Monleón, S. & Plomin, R. (2002) Evidence for general cognitive ability (g) in heterogenous stock mice and an analysis of potential confounds. *Genes, Brain Behav* **1**, 88–95.

Hascoët, M., Bourin, M. & Nic Dhonnchadha, B.Á. (2001) The mouse light-dark paradigm. A review. *Prog Neuro-Psychopharmacol Biol Psychiat* **25**, 141–166.

Hebb, D.O. & Williams, K. (1946) A method of rating animal intelligence. *J Genetic Psychol* **34**, 59–65.

Hulse, S.H. (1993) The present status of animal cognition: An introduction. *Psychol Sci* **4**, 154–155.

Humphreys, L.G. (1987) Psychometric considerations in the evaluation of intraspecies differences in intelligence. *Behav Brain Sci* **10**, 668–669.

Jensen, A. (1998) *The G Factor: the Science of Mental Ability*. Praeger, Westport.

Jensen, A.R. (1987) The *g* beyond factor analysis. In Ronning, R.R., Glover, J.A., Conoley J.C. & Witt, J.C. (eds), *Brain Mechanisms.* Annals of the New York Academy of Sciences, New York, pp. 87–142.

Jensen, A.R. & Weng, L.J. (1994) What is a good g? *Intelligence* **3**, 231–258.

Jensen, C., Schmitt, J.C., Scheirer, C.J. & Cochrane, T.L. (1978) Factor analysis of active avoidance and operant discrimination learning in mice. *Multivariate Behav Res* **13**, 45–61.

Locurto, C. (1997) On the comparative generality of g. In Tomic, W. & Kigma, J. (eds), *Advances in Cognition and Education, Vol. 4: Reflections on the Concept of Intelligence.* JAI Press, Greenwich, pp. 79–100.

Locurto, C. & Scanlon, C. (1998) Individual differences and a spatial learning factor in two strains of mice. *J Comp Psych* **112**, 344–352.

MacKintosh, N.J. (1998) *IQ and Human Intelligence*. Oxford University Press, Oxford.

Macphail, E.M. (1982) *Brain and Intelligence in Vertebrates*. Clarendon Press, Oxford.

Mclearn, G.E., Wilson, J.R. & Meredith, W. (1970) The use of isogenic and heterogenic mouse stocks in behavioral research. In Lindzey, G. & Thiessen, D. (eds) *Contributions to Behavior Genetic Analysis: the Mouse as a Prototype*. Appleton-Century-Crofts, New York, pp. 3–22.

Miller, L.T. & Vernon, P.A. (1992) The general factor in short-term memory, intelligence, and reaction time. *Intelligence* **16**, 5–30.

Mizumori, S.J.Y., Rosenzweig, M.R. & Kermisch, M.G. (1982) Failure of mice to demonstrate spatial memory in the radial maze. *Behav Neural Biol* **35**, 33–45.

Murphy, K.R. & Davidshofer, C.O. (1988) *Psychological Testing: Principles and Applications.* Prentice Hall, Englewood Cliffs.

Nunnally, J. (1967) *Psychometric Theory*. McGraw-Hill, New York.

Olton, D.S., Becker, J.T. & Handelmann, G.E. (1979) Hippocampus, space, and memory. *Behav Brain Sci* **2**, 313–365.

Olton, D.S. & Samuelson, R.J. (1976) Remembrance of places passed: Spatial memory in rats. *J Exp Psychol Anim Behav Process* **2**, 97–114.

Rabinovitch, M.S. & Rosvold, H.E. (1951) A closed-field intelligence test for rats. *Can J Psychol* **5**, 122–128.

Roberts, W.A. (1998) *Principles of Animal Cognition*. McGraw-Hill, Boston.

Spearman, C. (1904) ‘General intelligence’ objectively determined and measured. *Am J Psychol* **15**, 201–293.

Terrace, H.S. (1993) The phylogeny and ontogeny of serial memory: List learning by pigeons and monkeys. *Psychol Sci* **184**, 162–169.

Thompson, R., Crinella, F.M. & Yu, J. (1990) *Brain Mechanisms in Problem Solving and Intelligence: a Lesion Survey of the Rat Brain*. Plenum, New York.

Thurstone, L.L. (1934) The Vectors of Mind. *Psychol Rev* **41**, 1–32.

Thurstone, L.L. (1938) *Primary Mental Abilities*. Chicago University Press, Chicago.

Tolman, E.C. (1932/1967) *Purposive Behavior in Animals and Men*. Appleton-Century, Oxford.

Weiskrantz, L., ed. (1985) *Animal Intelligence*. Clarendon Press, Oxford.

Acknowledgments

The authors wish to thank Matthew Bletsch, Sarah Bolduc, Matthew Dooley, Gina DiPalma, and Meghan Sinton for their considerable help in the execution of this study. This research was funded by National Institute of Mental Health Grant R15MH59997–01A1 to Charles Locurto.

Appendix

Table A1: Correlation matrices for univariate measures and control procedures

	1	2	3	4	5	6	7	8	9	10	11
ETS											
1. Detour	1.00										
2. RAM	0.18	1.00									
3. 4 × 4	0.13	0.28^{**}	1.00								
4. HW	0.17	0.05	0.01	1.00							
5. Place	0.13	0.09	−0.11	0.18	1.00						
6. VNMTS	0.03	−0.08	0.07	−0.21	−0.12	1.00					
7. Land Activity	0.14	−0.05	-0.22^{*}	0.03	−0.18	0.04	1.00				
8. Thigmotaxis	−0.04	0.17	0.06	0.01	−0.06	−0.15	−0.21	1.00			
9. Light/Dark	0.03	0.08	−0.02	0.17	−0.04	−0.04	0.05	0.12	1.00		
10. LD Transitions	−0.06	0.08	0.18	0.12	0.06	−0.04	0.32^{*}	-0.25^{*}	−0.12	1.00	
11. Water Activity	0.07	0.07	−0.06	−0.01	0.07	0.06	0.06	0.18	−0.03	0.06	1.00
Errors											
1. Detour	1.00										
2. RAM	0.14	1.00									
3. 4 × 4	0.30^{**}	0.10	1.00								
4. HW	0.14	0.43^{**}	−0.13	1.00							
5. Place	0.27^{*}	0.30^{**}	0.08	0.67^{**}	1.00						
6. VNMTS	0.18	−0.10	0.21^{*}	−0.15	−0.19	1.00					
7. Land Activity	−0.09	−0.06	0.18	−0.12	0.01	−0.01	1.00				
8. Thigmotaxis	0.10	0.02	0.01	0.03	−0.02	−0.02	−0.21	1.00			
9. Light/Dark	−0.08	0.02	0.07	0.01	0.09	−0.02	0.05	0.12	1.00		
10. LD Transitions	−0.03	−0.06	0.16	0.00	0.04	0.23^{*}	0.32^{*}	-0.25^{*}	−0.12	1.00	
11. Water Activity	0.01	−0.17	0.25^{*}	−0.02	0.11	-0.23^{*}	0.06	0.18	−0.03	0.06	1.00
Latency											
1. Detour	1.00										
2. RAM	0.28^{*}	1.00									
3. 4 × 4	0.10	0.03	1.00								
4. HW	0.03	0.16	0.07	1.00							
5. Place	0.04	0.10	0.04	0.59^{**}	1.00						
6. VNMTS	−0.05	0.00	0.02	0.58^{**}	0.56^{**}	1.00					
7. Land Activity	−0.05	0.02	0.09	−0.08	0.13	−0.03	1.00				
8. Thigmotaxis	0.16	0.01	−0.01	-0.25^{*}	−0.09	−0.09	−0.21	1.00			
9. Lignt/Dark	−0.18	−0.03	−0.05	−0.08	0.25^{*}	0.15	0.05	0.12	1.00		
10. LD Transitions	0.06	0.05	0.23^{*}	−0.07	−0.03	−0.12	0.32^{*}	-0.25^{*}	−0.12	1.00	
11. Water Activity	−0.08	−0.03	−0.16	-0.26^{**}	−0.15	−0.11	0.06	0.18	−0.03	0.06	1.00

$^{*}P < 0.05$; $^{**}P < 0.01$.

Table A2: Rotated component (C) loadings, eigenvalues, and percent variance accounted for in univariate measures

	ETS				Errors			Latency		
	C1	C2	C3	C4	C1	C2	C3	C1	C2	C3
Task										
Detour	0.28	0.49	0.49	0.04	0.37	0.52	0.08	−0.04	0.79	−0.11
RAM	0.13	0.71	−0.06	−0.35	0.71	0.00	−0.22	0.09	0.80	0.02
4 × 4	−0.20	0.79	−0.09	0.19	0.00	0.88	−0.24	0.05	0.09	−0.62
HW	0.69	0.09	0.24	0.20	0.85	−0.08	0.12	0.84	0.09	−0.21
Place	0.63	−0.01	0.02	−0.11	0.78	0.19	0.22	0.82	0.10	0.14
Visual NMTS	−0.63	0.08	0.37	0.08	−0.23	0.14	−0.77	0.83	−0.08	0.10
Activity	−0.05	−0.15	0.81	−0.08	−0.04	0.07	0.53	0.00	0.02	0.82
Stress	0.01	0.03	0.06	0.92	−0.19	0.59	0.42	−0.37	0.01	0.22
Eigenvalue	1.41	1.39	1.10	1.07	2.10	1.45	1.22	2.24	1.30	1.18
% Variance	17.61	17.45	13.80	13.40	25.78	18.15	15.30	27.95	16.22	14.79

Table A3: Component correlations (oblique rotation)

	1	2	3	4
ETS				
1. C1	1.00			
2. C2	0.06	1.00		
3. C3	0.05	0.06	1.00	
4. C4	−0.08	0.00	0.05	1.00
Errors				
1. C1	1.00			
2. C2	0.04	1.00		
3. C3	0.04	0.09	1.00	
Latency				
1. C1	1.00			
2. C2	0.06	1.00		
3. C3	0.02	0.08	1.00	