Published in final edited form as:

Horm Behav. 2007 April ; 51(4): 496–507.

Cognitive Performance in Rhesus Monkeys Varies by Sex and Prenatal Androgen Exposure

Rebecca A. Herman¹ and Kim Wallen

Department of Psychology and Yerkes National Primate Research Center, Emory University, Atlanta, Georgia

Abstract

Men and women differ on performance and strategy on several spatial tasks. Rodents display similar sex differences, and manipulations of early hormone exposure alter the direction of these differences. However, most cognitive testing of nonhuman primates has utilized sample sizes too small to investigate sexually-differentiated behaviors. This study presents an investigation of sex differences and the effects of prenatal androgen on spatial memory and strategy use in rhesus monkeys. Monkeys prenatally exposed to vehicle, testosterone, or the androgen receptor blocker flutamide performed a search task in which 5 of 12 goal boxes contained food rewards. Spatial consistency and the presence of local landmarks were varied. Performance when both spatial and marker cues were available did not differ by sex or prenatal treatment. Contrary to predictions, females easily solved the task when local markers were removed, and their performance outscored males. Although eliminating spatial consistency and requiring subjects to use local markers impaired performance by all monkeys, females continued to locate correct goal boxes at higher than chance levels and scored better than males. Blocking prenatal androgen exposure in males improved use of local markers. These findings suggest that the tendency to attend to landmarks and to use them in solving spatial problems is typical of females across many species, including rodents, humans, and rhesus monkeys. In rhesus monkeys and rodents, developmental androgen eliminates this specialization. However, these results are the only known example of better performance of females than males when salient markers are removed.

Keywords

rhesus monkey; spatial cognition; sex differences; testosterone; flutamide; prenatal; landmarks

On a variety of spatially-demanding tasks, men outperform women. Tests of mental rotation, the ability to imagine the rotation of a two- or three-dimensional object rapidly, provide the largest and most reliable sex differences (Linn and Petersen, 1985), but performance differences exist in other spatial abilities also, including the ability to quickly and accurately learn a described route (Galea and Kimura, 1993; Ward et al., 1986). Not only do differences exist in overall performance, but men and women tend to use different strategies in map-learning and navigation. Men use cardinal directions and mileage estimates more, whereas women rely upon landmarks more (Galea and Kimura, 1993; Lawton, 1994, 1996; Saucier et al., 2002; Ward et al., 1986). Choice of strategy may be the important difference as route

¹To whom correspondence should be addressed: Department of Psychology Emory University 532 N. Kilgo Circle Atlanta, Georgia 30322 E-mail: rebecca.herman@gmail.com Phone: 678-637-1357 Fax: 404-727-0372

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

memory tended to correlate with cardinal and distance knowledge but not landmark recall in one study (Galea and Kimura, 1993). Men are also more accurate at pointing in the direction of a well-known but unseen landmark (Lawton, 1994, 1996). Similarly, Holding and Holding (1989) found that men were better at choosing the angle between two locations on a route they had just viewed on slides, suggesting perhaps the formation of a more accurate cognitive map.

Studies of rats have identified somewhat similar sex differences in both performance and strategy. Males often prove to be better than females at remembering the location of and traveling directly to a submerged platform in the Morris Water Maze (MWM; Isgor and Sengelaub, 1998; Roof and Havens, 1992). Kanit and colleagues (1998) revealed a sex difference in strategy on a modified water maze task in which the platform was most often submerged but was sometimes visible. Male rats relied solely upon spatial location to find the platform whereas female rats used a strategy of first looking for the platform, and if they failed to see it, swimming to the remembered location. In this case, the “male” strategy is more efficient when the task remains unchanging, but the “female” strategy is more adaptive to changing conditions.

On a radial arm maze (RAM) with a constant subset of arms baited with food, gonadectomized adult male rats chose arms more accurately than did gonadectomized adult females (Williams et al., 1990). Manipulating the cues available to the rats indicated that male performance was impaired by changes to the room geometry, regardless of the presence, absence, or rearrangement of landmarks. Females, however, were unaffected by alteration of the geometry as long as landmarks were unaltered and were similarly unaffected by removal of landmarks if geometry was constant. If landmarks were rearranged, thereby changing their relationship with the geometry of the room, females were impaired. In other words, female rats used both geometry and landmarks, independently or in combination, to locate the food items. The authors suggest that male rats' reliance on a single type of cue may explain their faster acquisition of the task. A related study with human subjects that used a virtual water maze and manipulated available cues found that men used geometry or landmarks to locate the hidden platform, whereas females relied upon landmarks only (Sandstrom et al., 1998). The use of single or multiple cues by sex differs in these studies, but in both cases, females used landmarks and males used room geometry to navigate.

Cognitive sex differences in rodents can be manipulated by altering sex hormone exposure during a period of neural organization. Early treatment (prenatally or neonatally) of female rats with testosterone (T) or its derivative dihydrotestosterone (DHT) improves adult female performance on the MWM (Isgor and Sengelaub, 1998; Roof, 1993; Roof and Havens, 1992) and RAM (Roof, 1993), whereas neonatal castration of males or treatment with antiandrogens hinders adult performance on the MWM (Isgor and Sengelaub, 1998). Strategy use is also affected by hormonal conditions during organization, such that males castrated neonatally performed just like control females in the RAM study by Williams and colleagues (1990). Castrated males made more errors during acquisition of the partially-baited maze and used both landmarks and geometrical cues. Female rats treated with estradiol benzoate, which, like neonatal T, can masculinize behavior, were like control males and unlike control females on the same tasks.

In humans, however, such experimental manipulations are not possible, and socialization processes have been proposed to explain human sex differences in spatial abilities, suggesting that differential experiences through play and encouragement of different educational pursuits in boys and girls result in males outperforming women and girls on some spatial tasks. Significant correlations are found between spatially-demanding activities and performance on cognitive measures (Newcombe et al., 1983), but the causal relationship between experience and spatial performance is unknown. Attempts to study the role of early hormone exposure

have largely been limited to clinical populations. Girls with congenital adrenal hyperplasia (CAH) are exposed to excess prenatal and neonatal androgens due to hypersecretion of adrenal androgens as a result of errors in steroid biosynthetic enzymes required for glucocorticoid production. Studies of spatial abilities in girls with CAH have reported mixed results, but the most careful studies have found that CAH girls and women have better spatial abilities than their unaffected female relatives (reviewed in Berenbaum, 2001). Resnick et al. (1986) not only found that females with CAH excelled at sexually-differentiated spatial skills, but that their increased participation in spatial manipulation activities as children did not account for the effect.

Cognitive studies in nonhuman primates could help resolve the question of prenatal testosterone's role in cognitive sex differences. However, few studies have investigated sex differences in cognition in adult, nonhuman primates. Most studies of cognition in monkeys and apes simply have not sample sizes large enough to consider sex as a variable. McDowell et al. (1960) found that young female rhesus monkeys outperformed young males on a spatial delayed response task, but they attributed this difference to greater concentration by the females and more distractibility in the males. More recently, young adult male rhesus performed better than same-aged females at a Delayed Recognition Span Test in which disks are added one by one on top of a matrix of wells, and the subject must successively identify each new disk location to reveal a food reward (Lacreuse et al., 1999). Some nonhuman primate studies, while lacking samples large enough to assess sex differences, have considered spatial cognition in general. For example, wild capuchins proved to have good memory by quickly learning which of 13 locations consistently contained bananas (Garber and Paciulli, 1997). Interestingly, when the locations of the bananas were no longer constant but instead were marked by the presence of a yellow block on the baited platform, the capuchins performed worse, indicating greater ease in using spatial cues than local markers. Other studies have demonstrated a tendency to follow efficient pathways between food locations (Cramer and Gallistel, 1997; Menzel, 1973), indicating good navigation skills and the likely formation of a cognitive map.

In the work presented here, young adult rhesus monkeys navigated a 4.9m × 4.9m open area to locate highly-valued food items in goal boxes. The consistency of the food locations and the presence of colored markers on baited goal boxes were manipulated to assess subjects' use of spatial arrangement and local markers. This design is similar to those of other researchers who have used multiple locations within an open area to study spatial memory in nonprimate species (in birds: Balda and Kamil, 1988; Kamil et al., 1994; Olson et al., 1993; in mice: Dell'Omo et al., 2000) and closely resembles a paradigm recently used to study spatial memory performance in squirrel monkeys (Ludvig et al., 2003). The test area and the types of cues available differ however from those used in most cognitive studies in rodents and virtual studies in humans. Subjects include males, females, animals of both sexes that received an androgen receptor blocker (flutamide) prenatally, and females who received testosterone prenatally. These subject groups allow for investigation of sex differences in spatial memory and cue usage and for the effects of prenatal androgen manipulations on performance

Methods

Subjects

Subjects were the 51 offspring of female rhesus monkeys who received experimental prenatal treatments at the Field Station of the Yerkes National Primate Research Center. Throughout the study, subjects remained in their natal multi-male, multi-female, stable social groups of 60-150 animals. Subjects were created in two cohorts over two consecutive birth seasons (March - June of each year) and were 5 to 8 years old at the time of testing. All research was approved by the Institutional Animal Care and Use Committee and developed in accordance with the *NIH Guide for the Care and Use of Laboratory Animals*.

The method used to time conception has been described in detail elsewhere (Zehr et al., 2000). Briefly, it involved observations of female sexual initiation and verification of pregnancy and developmental age using ultrasonic visualization. The two estimates of conception date were highly correlated ($r = 0.98$). Pregnant females received either supplementary androgen (20 mg/week of testosterone enanthate im at a concentration of 100 mg/ml dissolved in sesame oil), the androgen-receptor blocker flutamide (30 mg/kg twice daily of flutamide im dissolved in dimethyl sulfoxide (DMSO), at a concentration of 500 mg/ml), or twice daily DMSO vehicle injections (all reagents from Sigma Chemical Company, St. Louis, MO). Treatment was started either early (estimated gestational day (GD) 40 in cohort 1 or GD 35 in cohort 2) or late (GD 115 in cohort 1 or GD 110 in cohort 2) in gestation (~170 days) and continued for 30 (cohort 1) or 35 (cohort 2) days. Vehicle control pregnancies early and late in gestation are combined for this study, and an additional two males and three females whose mothers did not receive vehicle treatment were added to the control groups early in development to increase sample size. Except for the prenatal vehicle treatments and associated handling, these five unmanipulated subjects underwent the same procedures and received comparable handling throughout their lives. Treatments produced six experimental and two control groups of subjects as shown in Table 1. Androgen-treated male offspring were not included in this study as a result of subject loss over the 8 years of this extended study.

All subjects have been studied throughout infancy, the juvenile period, and puberty, and subtle but significant effects of the treatments have been seen in their morphology and endocrinology (Herman et al., 2000), developmental timing (Herman et al., 2006), vocal communication (Tomaszycki et al., 2001), and social behavior (Herman et al., 2003; Wallen, 2005).

For all habituation, training, and testing, subjects were temporarily removed from their social groups using procedures to which all subjects were already fully accustomed. All females tested in the first three months of testing performed trials with their infant offspring present on their ventrum. In the later months, all infants were removed from their mothers and returned to the social groups before testing each day. Observations during and after testing indicated that neither mothers nor infants were distressed by the temporary separation; mothers continued to perform the task willingly, and infants left within their social groups were observed to show normal behavior and no distress calls. In total, 11 females were tested with infants, 12 females had infants removed, and 6 had no infant.

The Testing Facility

The testing facility was an approximately 4.9m L $\times$ 4.9m W $\times$ 2.4m H area with chain link walls and roof and a concrete floor. Animals entered the facility from a vestibule on the north side in which the researcher recorded the animal's activity, and subjects exited the testing area into a capture area in the northwest corner of the facility. Initially, subjects had to be ushered into the capture area at the end of a trial, but they quickly learned to run through the door to the area as soon as it was opened, signaling the end of a trial. A carport surrounded the facility, providing protection from rain and blocking visual access of the road and monkey group to the west. A chain link fence with vertical slats along the northeast side obscured visual access to the nearby monkey group. Blocking visual access of surrounding monkey groups limited, but could not completely eliminate, disturbances, such as the sound of monkey vocalizations from nearby groups. However, the testing facility was sufficiently distant from the subjects' home groups to prevent any visual or auditory access to them. Finally, the area surrounding the facility remained rich with cues visible to the subjects at all times, including the vestibule to the north, an extension of the concrete pad to the east, and a road and a wooded area to the south.

Twelve goal boxes were attached to the chain-link along the south, east, and west sides of the testing facility in an irregular pattern. The goal boxes consisted of 5-inch long, 2-inch diameter PVC Tees with the "T" extension removed, producing a round opening in the center of the

pipe. Both ends of the pipe were closed with PVC plugs. The locations of the 12 goal boxes were constant throughout testing. To bait a goal box, food items were placed within the tube and were approximately two inches below the base of the opening. When a colored marker was required, a blue 4-inch diameter plastic circle was attached to the bottom PVC plug.

Habituation and Training

All subjects received habituation trials within the testing facility in order to familiarize them to the facility, the presence of food treats, and observation. Eight food items (plain M&M's™ grapes, raisins, or peanuts) were distributed around the floor of the facility in a nonsystematic, varying pattern with no goal boxes present. Subjects received daily (4-5 times a week) habituation trials, lasting ten minutes or until they had eaten all the food items. Individual subjects completed habituation once they ate seven or eight treats within ten minutes on three out of four consecutive training days. Five subjects (Table 2, 2 EFM, 1 LFM, 2 LFF) never passed the criterion for habituation, after more than 30 attempts. Nevertheless, these subjects were allowed to proceed to the training portion of testing. The other subjects required on average 6.7 ± 1.0 days to pass habituation.

Unlike habituation, training trials included the presence of baited goal boxes. The 12 goal boxes were placed in their set locations on the fence walls, and each of the 12 contained 3 plain M&M's™ 3 raisins, and 2 peanuts. (Pilot testing of four animals not included in the study suggested that baiting goal boxes with only a single food treat did not provide enough incentive for the animals to perform the search task.) On initial training days, additional baited goal boxes and extra food items were placed on the floor of the test area. An animal completed training after at least five training trials and once he or she investigated five unique goal boxes within a 5-minute trial. For animals that did not easily transfer from eating out of the goal boxes on the floor to visiting the goal boxes on the fence, various techniques encouraged greater exploration. As a subject completed training, he or she immediately began Dual Cue Acquisition on the following day of testing.

Of the five animals that did not successfully complete habituation, two (1 LFF (Se6), 1 EFM (Bh6)) also failed to complete training in over 30 trials and were therefore dropped from the study (Table 2). A third (LFM (Nv5)) completed training but was subsequently dropped during testing when he consistently failed to visit goal boxes. The other two subjects that did not complete habituation nevertheless trained successfully, although one (EFM, Js5, see below) would eventually be dropped for failure to learn the task. In addition, two other successfully habituated LFM's did not complete training in over 40 attempts and were dropped from the study (Cd6 and Ks5). Dropping of these 5 subjects resulted in sample sizes shown in Table 1. With only 2 LFM's remaining, statistical comparisons involving this treatment group were not possible, and these two males are therefore discussed qualitatively or only as part of the larger male group throughout the remainder of this paper.

Testing Protocol

Order of testing throughout the test period (October 2003 – August 2004) was balanced across prenatal treatments within both sexes. However, all females were tested first and males second. This allowed all males to be tested within the nonbreeding season, limiting within sex variation due to large seasonal fluctuations in both endocrinology and social behavior. Because sex differences in other species are larger under breeding conditions than nonbreeding conditions (Galea et al., 1994), testing males in the nonbreeding season, when testosterone levels are at their lowest, provides a conservative test for cognitive sex differences. During testing, female subjects underwent ovarian cycles, conception, and pregnancy. Periodic blood sampling was conducted to permit analysis of effects of variation in circulating ovarian hormones and

testosterone in both males and females. The results from these analyses will be presented separately.

Subjects underwent trials five days a week (most often Monday – Friday), once per day, and once a subject began testing, no more than three days passed without a test trial. During each test trial, a human observer recorded the subject's behavior and goal box visits with a PDA (Palm IIIxe, Palm Corp., Milpitas, CA) with attached keyboard (GoType!Pro, LandWare, Oradell, NJ) using a program (HandObs, CBN, Atlanta, GA) that applies an automatic time stamp after each data line. Visits were defined as any time a subject reached into or clearly looked inside a goal box. One researcher (RH) observed ninety-five percent of the test trials. Trials continued for 5 minutes or until a subject had visited all five correct goal boxes. If a subject failed to visit any goal boxes during the 5-minute trial (correct or incorrect), that trial was dropped and the same trial number was repeated on the following test day. Only five subjects had more than five such trials throughout the testing period (Table 2), and these five subjects were nevertheless able to complete testing successfully.

Trials were recorded by two videocameras placed outside the testing facility 90 degrees from each other, thereby permitting calculation of the subject's location in three dimensions throughout the trial (see Khan et al., 2005). Further results from this videotracking will be presented separately.

Animals began testing with Dual Cue Acquisition (DCA). After 24 acquisition trials or upon reaching criterion, half of the subjects, balanced within sex and prenatal treatment group, began the Spatial Task (ST) whereas the other half began the Marker Task (MT). Four repeat trials of the Dual Cue Acquisition task (rDCA) followed as a “reminder” for the subjects. Finally, subjects completed the remaining task (Spatial or Marker).

Dual Cue Acquisition (DCA)—For each DCA trial, the same five goal boxes displayed colored plastic markers and contained 3 plain M&M's™, 3 raisins, and 2 peanuts each. These five locations were spread across the 12 possible locations and remained the same for all subjects. Testing continued for 24 trials or until subjects reached criterion. To reach criterion, subjects visited four correct goal boxes in their first five visits for two consecutive trial days. Because subjects could utilize either the blue plastic markers or the constant spatial locations to find food, this was referred to as a “dual cue” task. Olfactory cues were not expected to appreciably assist the subjects. Although capable of excellent olfactory discrimination, macaques do not likely preferentially use olfactory cues (Hubener and Laska, 2001). Moreover, on any given day of testing, various subjects were in different stages of testing, including training. Thus, all goal boxes likely contained food at some point, and food odors were unlikely to have accumulated differentially within any given goal box. Additionally, the food treats used were not highly aromatic, and because the facility is outdoors and in the vicinity of other animal areas, it is in an olfactory-rich environment, which should further reduce any possibility of using olfactory discrimination to solve the task.

Spatial Task (ST)—Within each prenatal treatment group, subjects were alternatively assigned to complete the ST first or the MT first after they completed DCA. For the ST, the five blue plastic markers were removed from the goal boxes, but the food remained in the same locations. Other procedures remained as described for DCA. Each subject received four ST trials. Our Spatial Task differs from the “geometric” tasks in other studies discussed above (Sandstrom et al., 1998; Williams et al., 1990) in that although the most informative and salient landmarks are removed, many distal landmarks remain. The resulting task is clearly spatial in nature, but does not necessitate dependence on shape of the environment. Relative location of goal boxes from landmarks, including the start box and the capture area of the testing facility, could also be used.

Marker Task (MT)—The four trials of the MT additionally probed the subjects' use of specific cues. Performance on the task assessed reliance on the direct identification of goal locations provided by the colored markers versus utilization of spatial information. For the MT, the blue plastic markers were moved each day to five different locations, and the food moved with the markers. On day 1, the five locations were chosen so that there was no overlap between the previously correct five location and the new locations. On days 2-4, locations were chosen randomly with the caveats that no more than two locations from the previous day or two locations from DCA could be included. The correct locations for each trial day were the same across subjects.

Scoring of Test Trials

The first visit within a trial to a goal box containing food was recorded as a correct visit. A return visit to this goal box during the same trial, traditionally referred to as a working memory error, could in this case reflect either a memory error or a return to finish remaining food items. Therefore, these visits were not recorded in visit totals or in analyses. In practice, subjects tended to eat all the food in a goal box when they first visited it, and return visits to correct goal boxes were rare. Working memory errors were instead defined only as a repeat visit to an incorrect goal box. The first visit within a trial to an incorrect goal box produced a reference error.

A performance score was calculated as the primary measure of performance. Determination of the performance scores first required determination of a difference score, given by (correct visits – reference errors). With 5 of 12 goal boxes initially correct, the probability of a given difference score depends on the total number of goal boxes visited. The difference score that would be obtained by chance was therefore calculated for each possible number of total visits from 1 to 12 and was called the “chance score.” Subtracting the chance score from the difference score produced the performance score. For example, if a subject visited 7 goal boxes, 4 of which were correct, its difference score would be 1 (4 minus 3). The chance score for 7 visits is -1.189 . Thus, the resulting performance score would be 2.189 (1 minus -1.189). A performance score of zero indicates chance performance and positive scores indicate better than chance performance.

This performance score has an advantage over measures used in other studies of radial arm mazes and their analogues, such as the total number of choices required to find all food items in that it allows for scoring trials in which subjects did not locate all of the food items. Moreover, this performance score provides a control for the total number of visits, which differed between subjects, possibly due to differences in motivation, anxiety, or personality.

Working memory errors proved to be very rare. Therefore, they are considered separately within each part of the study and are not included in the performance scores.

Statistical Analysis

On all measures, within all tasks, sex differences were first considered. Because previous studies with this subject group have indicated that prenatal treatment effects would likely be small in magnitude compared to sex differences, sex differences were first analyzed by comparing all males ($n = 16$) and all females ($n = 29$). Doing so increases samples sizes and therefore statistical power but should, if anything, minimize sex differences because of the groups included. The presence of infants affected performance of some females (see below), and therefore, sex differences also were analyzed by comparing all males and females without interfering infants ($n = 20$). These results are described only when they differ from those found with the inclusion of all 29 females. Finally, analyses compared control males and control females. One-way ANOVAs, with least significant differences *post-hoc* analyses when

necessary, and independent *t*-tests indicated the presence of any prenatal treatment effects within females and within males (controls and EFM), respectively. Because data were available from only two LFM, the performance of these males could only be discussed qualitatively when considering treatment effects. Because many subjects scored no working memory errors, nonparametric analyses were used in comparing working memory performance.

Performance of subjects completing the MT immediately after DCA did not differ from subjects completing the ST first, and data from both groups were combined for all analyses.

All analyses were performed with SPSS 11.5 or SPSS 12.0 for Windows. In all analyses, an alpha level of 0.05 was used. Probability values between 0.05 and 0.10 are discussed as indicative of trends.

Results

Problem subjects

Table 2 lists the treatments of all subjects that had difficulties during the testing procedures, including numerous trials with zero visits, failures to successfully complete testing, and failures to habituate and train. One male subject, an EFM (Js5), completed testing trials but never showed improvement. On the last four DCA trials, he visited one or two goal boxes only and his performance score was 0.25 ± 0.41 , with zero indicating chance performance. Because he consistently failed to fully participate in the task, this subject was dropped from all analyses of performance, leaving 5 EFM subjects (Table 1).

Across subjects, an association between flutamide treatment and difficulties in training and testing emerged. Analysis indicated significant differences in the proportion of “difficult” subjects among treatment groups: (androgen treatment: 1 of 11 subjects with difficulties, flutamide treatment: 10 of 25 subjects with difficulties, and control treatment: 0 of 15 subjects with difficulties; $G = 12.83$, $df = 2$, $p = 0.002$). Males also tended to have more difficulties than did females (7 of 21 males vs. 4 of 30 females; $G = 2.89$, $df = 2$, $p = 0.09$).

Interference by infants

Eleven females were tested with their 5-9-month-old infants present, 12 females had infants removed, and 6 had no infant. Infants sometimes interfered with testing by moving away from their mothers and visiting a correct goal box before their mother did so. Of the 11 females tested with infants present, 9 had infants that interfered in this manner on at least 10 percent of test trials. On a given trial, if an infant visited a correct goal box before his or her mother, the mother visited that same goal box during the same trial with a 34% likelihood. In contrast, the probability of these same mothers visiting a correct goal box not first visited by their infants was 62%. Thus, mothers were not following their infants to correct goal boxes but instead may have been hindered by their infants. The 9 females with interfering infants performed somewhat worse throughout testing than did the other 20 females. Their performance scores on the MT were significantly worse (with infants: 0.12 ± 0.27 , without infants: 0.75 ± 0.14 , $t(27) = 2.22$, $p = 0.04$) and they showed a significantly larger initial drop in performance score on the ST than did females without interfering infants (with infants: -1.55 ± 0.43 , without infants: -0.18 ± 0.37 , $t(27) = 2.19$, $p = 0.04$). There were also trends for the nine females to perform worse on the final 4 trials of DCA (with infants: 2.93 ± 0.30 , without infants: 3.57 ± 0.20 , $t(27) = 1.80$, $p = 0.08$). These differences are likely due to the reluctance of the females to visit a correct goal box once the infant has already found the food within, but could also be explained by greater distraction (and thus poorer learning) because of the presence of the wandering infant.

The presence of interfering infants did not differ across prenatal treatment groups (3 EAFs, 1 LAF, 2 EFFs, 1 LFF, 2 controls). However, whenever the presence of sex differences differed when the nine females with interfering infants were excluded, the results for comparisons both with and without these females are discussed.

Testing Performance

Dual Cue Acquisition (DCA)—Thirty-one of 45 subjects (69%) reached criterion in 24 or less trials with ten trials being the smallest number of trials required (an EFM). One other male reached criterion in 11 trials (a control male), and three females did so in 12 trials (1 control female, 1 EFF, 1 LAF). In total, 20 of 29 females and 11 of 16 males reached criterion in 24 or less trials. A Kaplan-Meier Survival Analysis compared the number of trials required across groups, taking into consideration the number of subjects that never reached criterion. Males reached criterion in an average of 17 trials (confidence interval (CI) = (15, 20)) whereas the 29 females required 19 trials (CI = (17, 21)). A log rank factor level comparison indicated no significant difference between the sexes. The log rank test also indicated no difference in the time to criterion for male and female controls, female treatment groups, or male treatment groups. The 31% of subjects that did not reach criterion still clearly demonstrated learning of the task. Their performance scores over the last four trials of DCA were well above chance (mean = 2.50 ± 0.20 , one-sample *t*-test vs. zero: $t(13) = 12.4$, $p < 0.001$).

Because subjects experienced different numbers of Dual Cue Acquisition (DCA) trials, average performance measures across all of the trials are not directly comparable. However, analysis of scores on the last four DCA trials should indicate whether animals reached the same level of performance on the task. On these trials performance scores did not differ between males and all females. However, eliminating the nine females with interfering infants produced a significant difference in performance scores (Figure 1; $t(34) = 2.20$, $p = 0.04$), with females outperforming males. Control females and males did not differ on the last block of DCA trials. Nor did prenatal treatment affect scores. Additionally, the single best performance score across DCA for each individual did not differ by sex (all males and females, all males and females without interfering infants, control males and females) or treatment (independent *t*-tests and one-way ANOVAs, $ps > 0.10$ for all; data not shown).

The possibility that subjects relied upon response patterns or algorithms instead of spatial or marker memory to solve the task was considered by studying the order in which goal boxes were visited for each subject across the DCA trials. If a subject relied on such a motor strategy, the order of visits would necessarily remain constant. However, no subject consistently visited goal boxes in the same order. For one-third of subjects (30% of females, 40% of males), a given goal box was the first box visited for at least the last three trials. A single female (LAF) visited the same three goal boxes first, second, and third on each of her last three trials, but for other subjects goal boxes visited after the first varied from trial to trial. Only 1 of the 45 subjects (CM) visited goal boxes in an identical order on his last two trials, and this pattern was never observed on his earlier trials.

The variation in order of visits strongly suggests that the animals were not relying upon response strategies or memorized motor patterns to solve this task. The one-third of subjects whose first visit was to a consistent goal box likely relied upon a set pathway from the stable start location to the first visit, but clearly did not do so for the remaining visits. It seems likely that stable patterns of visits would have emerged if more DCA trials had been permitted each animal, as motor habits developed. Instead, setting the criterion level as we did meant very few perfect or near perfect trials were recorded, and the ability to probe which cues animals used in the learning process was preserved.

Working memory: Working memory errors (return visits within a trial to incorrect goal boxes already visited) were very rare. Fourteen subjects (8 females, 6 males) never made a working memory error during DCA. On average, subjects made 0.16 ± 0.03 working memory errors per DCA trial. Of all visits during DCA, $2.1 \pm 0.4\%$ were working memory errors. These values differed neither by sex nor treatment (nonparametric Mann-Whitney U and Kruskal-Wallis tests, $p > 0.10$ or all). Initially, working memory errors were more common as subjects learned the properties of the task. On the first day of DCA, just over one-third of subjects made at least one working memory error (16 of 45 subjects, 7 males and 9 females). There was no effect of sex or treatment on working memory error frequency. On the second day 12 subjects made such an error, and by the third, the number of individuals had decreased to five. The nonparametric Friedman test indicated a significant drop across the first three days in the percent of total visits that were working memory errors ($X^2(2) = 7.30, p = 0.03$).

Summary: Subjects clearly demonstrated acquisition of the task, and time required to reach criterion did not differ by sex or treatment. Females outscored males on later performance, but this was a minor difference that did not foretell quicker attainment of criterion performance, and the best performance by each subject was equal in males and females. Effects of prenatal androgen manipulations were not apparent during DCA

Spatial Task (ST)—Removal of the colored markers did not drastically affect performance across subjects (Figure 1). In comparison to performance across the previous four trials (average for each subjects' last 4 DCA trials or the four repeat DCA trials), performance score did not change with the four ST trials in males (before ST: 2.75 ± 0.18 ; ST: 2.82 ± 0.24 , paired t -test, $p > 0.10$). In females, scores averaged across the four-trial blocks improved with the removal of the colored markers (before ST: 3.06 ± 0.17 ; ST: 3.44 ± 0.22 , $t(28) = 2.06, p = 0.05$). By prenatal treatment group, this improvement was only apparent in LAFs (before ST: 2.92 ± 0.36 , ST: 3.68 ± 0.26 , $t(5) = 4.96, p = 0.004$). Comparing males and females revealed a trend for better female performance scores across the ST ($t(43) = 1.82, p = 0.08$), which became significant when females with interfering infants were dropped ($t(34) = 2.74, p = 0.01$). Performance of control males and females did not differ (CM - 3.06 ± 0.34 , CF - 3.41 ± 0.61). Within both sexes, prenatal treatments did not affect scores (data not shown).

A repeated measures ANOVA of performance scores with ST Trial as the repeated measure and sex as a between-subjects factor revealed a trend for a main effect of trial (Figure 1, $F(3, 129) = 2.63, p = 0.05$; trials 2 & 3 > trial 1), and a trend for a main effect of sex ($F(1, 43) = 3.31, p = 0.08$), but no interaction between sex and trial ($F(3, 129) = 1.16, p > 0.10$). Eliminating females with interfering infants from the analysis also eliminated the trial effect, but produced a significant main effect of sex ($F(1, 34) = 7.47, p = 0.01$). Repeated measures analysis of control males' and females' performance scores provided no significant main effects or interaction.

The lack of an impairment on the ST as compared to previous performance was surprising. Therefore, the first trial of the ST was analyzed independently for any indication of an initial drop in performance. Female performance scores dropped significantly less from the previous trial to the first ST trial than did the scores of males. This was true when comparing all males and females ($t(43) = 2.04, p = 0.047$), or only control males and females (Figure 2, $t(13) = 2.69, p = 0.02$). Not only did the females show a smaller drop in performance scores than did males, in males this impairment was significant (one sample t -test vs. zero, $t(15) = -4.90, p < 0.001$) but in females, the drop in score only showed a trend ($t(28) = -1.98, p = 0.06$), and was nonsignificant in females without interfering infants ($t(19) = -0.50, p > 0.10$). The magnitude of impairment did not differ by prenatal treatment in either sex. However, the drop in performance scores with the first trial of the ST was significant in LFFs (Figure 2, $t(5) = -3.03, p = 0.03$) but in no other female treatment group ($p > 0.10$ for all). Both EFM and control

males showed a significant drop in performance scores (Figure 2; EFM: $t(4) = -2.91, p = 0.04$; control males: $t(8) = -4.38, p = 0.002$), and the two LFMs averaged a similar drop.

Working memory: Working memory errors did not increase with the removal of the markers in the ST. In fact, there were only a total of 4 working memory errors in 3 females (1 LFF, 1 EFF, 1 control) across the four trials.

Summary: Subjects were minimally impaired by the removal of the color markers in the Spatial Task. However, male scores dropped significantly more than did female scores on the initial trial, and, although the male scores increased quickly, their overall performance on the Spatial Task remained worse than that of the females. Prenatal blockade of androgen did not alter male performance, but treatment effects were evident in some cases in the females. Only late in gestation androgen-treated females showed significant improvement from the previous block of trials to the Spatial Task trials whereas late-gestation flutamide treatment of females was associated with initial drops in performance.

Marker Task (MT)—Altering the parameters of the task such that local cue usage was required and spatial location was uninformative produced significant drops in performance for all subjects – males, females, controls of either sex, and all prenatal treatment groups (Figure 1). However, performance scores of females across the four trials of the MT remained higher than those of males (all males and females: $t(43) = 3.64, p = 0.001$; control males and control females: $t(13) = 2.63, p = 0.02$). Moreover, performance on the MT exceeded chance performance in females, ($t(28) = 4.00, p < 0.001$), but male performance did not differ from chance levels ($t(15) = 1.56, p > 0.10$).

Performance scores did not differ within female treatment groups (Figure 3), but within the males, flutamide treatment early in gestation resulted in significantly higher performance scores than in the control males (Figure 3; $t(12) = 2.64, p = 0.02$). The two LFMs had scores closer to those of the control males.

Comparing performance scores on the four trials of the MT to chance (a score of zero; Figure 3) indicated that only performance of LAFs ($t(5) = 3.52, p = 0.02$) and LFFs ($t(5) = 3.60, p = 0.02$) exceeded chance. Other groups of females and EFMs did not differ from chance ($p > 0.10$), but control males actually scored significantly worse than chance ($t(8) = -2.82, p = 0.02$).

Repeated measures ANOVAs of performance scores revealed significant improvement across the four trials of the MT within females (Figure 1, $F(3, 84) = 6.40, p = 0.001$) with scores on trials 2, 3, and 4 exceeding trial 1 scores. Including prenatal treatment as a between-subjects factor in the analysis produced no significant interactions with trial. Repeated measures ANOVAs of males indicated that, in contrast to females, male performance scores failed to improve across the four trials (Figure 1, $F(3, 45) = 1.70, p > 0.10$). Comparison of EFMs and control males revealed no significant interaction between treatment and trial.

The performance above chance of females suggests that female subjects may have retained some understanding of the relevance of the colored markers from Dual Cue Acquisition (DCA). However, their improvement across the trials of the MT also raises the possibility that they simply learned the significance of the markers within these four trials. Comparing their scores on the MT to their scores on the initial trials of DCA could help resolve this issue. Better scores on the MT than on the initial DCA trials, when both colored markers and spatial information were available, would suggest a carry over of knowledge from the DCA. The first trial of the MT is not included in this comparison, because, regardless of understanding of the relevance of colored markers, subjects could not know whether the spatial or the marker cues were informative on this first trial. Thus, the second to fourth trials of the MT were compared to the

second to fourth trials of DCA. For females, but not males, performance scores on these MT trials were significantly higher than scores on corresponding DCA trials (females DCA: 0.241 ± 0.15 , females MT: 0.874 ± 0.19 , $t(28) = -3.19$, $p = 0.003$; males DCA: 0.126 ± 0.14 , males MT: -0.047 ± 0.23 , $t(15) = 0.655$, $p > 0.10$). By prenatal treatment group, the difference on performance scores was significant for EAFs (DCA: -0.269 ± 0.25 , MT: 0.743 ± 0.27 , $t(4) = -3.90$, $p = 0.02$) and LFFs (DCA: -0.032 ± 0.31 , MT: 1.257 ± 0.32 , $t(5) = -4.80$, $p = 0.01$) only. Thus, females, particularly EAFs and LFFs, but not males, likely gained some understanding of importance of the colored markers during Dual Cue Acquisition.

Performance scores of control males not only failed to exceed chance levels but were significantly worse than chance. This surprising finding could be explained by *perseverative* behavior – a failure to stop making choices that were initially rewarded but are no longer. To judge whether subjects perseverated, each trial of the marker task was rescored as if it was a DCA trial, with the same five correct DCA goal box locations coded as correct for all MT trials. If these new scores held steady across the four trials of the MT, perseverative behavior would be indicated. In females, the resulting performance scores decreased across the four trials (Figure 4, $F(3, 84) = 3.72$, $p = 0.01$, trial 1 > trial 3). However, in males perseverative scores did not significantly change across the four trials ($F(3,45) = 1.10$, $p > 0.10$). Males continued to visit previously correct goal boxes for, on average, 3 of the first 5 visits. When repeating this analysis by treatment group, with their smaller sample sizes, the decrease in perseverative behavior was seen in LFFs only ($F(3,15) = 5.26$, $p = 0.01$). Within males, control males and EAFs did not differ from each other and neither showed a significant change in scores over the four trials. A two-factor ANOVA with both sex and scores as variables did not indicate an interaction between sex and trial however. The apparent increase in female scores on trial 4 (see Figure 4) may partially explain this negative finding; on trial 4 two previously correct choices were once again correct whereas only one such goal location was correct on trials 3 and 4. Thus, subjects correctly using the markers would receive higher perseverative scores.

Working memory: Errors of working memory, calculated as errors per trial or the proportion of all visits that were working memory errors, were significantly more common on the MT than on the previous block in all subjects (Wilcoxon matched pairs signed rank test, errors per trial: $Z = -5.25$, $p < 0.001$; proportion of visits as errors: $Z = -5.20$, $p < 0.001$), but neither sex nor treatment affected either measure of working memory during the MT (Mann-Whitney U or Kruskal-Wallis test, $p > 0.10$ for all). However, considering the initial MT trial only, a significantly greater proportion of visits by females were working memory errors than were in males (Mann-Whitney U test, $Z = -2.18$, $p = 0.03$), although this difference was reduced to a trend when females with interfering infants were eliminated ($Z = -1.99$, $p = 0.08$). Prenatal treatment did not affect working memory error frequency on trial 1 of the MT. Although a sex difference was evident, working memory errors remained rare; 25 subjects made no working memory errors on the first trial, and only seven subjects made more than one such error. Revisits to incorrect locations were overwhelmingly to locations that were correct during DCA (34 of 36 errors or 94%), but, given the design of the first trial of the marker test, 71% (5/7) of goal boxes that could produce working memory errors were correct during DCA.

Summary: When spatial location was made an unreliable cue and only colored markers designated correct locations, all subjects exhibited poor performance. However, male performance was significantly worse than female performance. Females improved across the four trials, performed above chance levels, and demonstrated some carry-over knowledge of the value of the colored markers. In contrast, males failed to improve and did not exceed chance performance, possibly because of a tendency to perseverate. The greater tendency of females to make working memory errors contrasts with their better overall performance on the task. Early in gestation treatment with flutamide produced better performance within males.

However, in females, it was late gestation flutamide treatment that was consistently associated with better performance. Surprisingly, androgen treatment of females, both early and late, also resulted in better performance on some measures.

Discussion

This study effectively demonstrates the feasibility of investigating sex differences in rhesus monkey cognition with the use of a search task within a large arena. These results provide evidence that male and female rhesus monkeys utilize environmental cues differently and that, in some cases, these differences are affected by prenatal androgen exposure. Moreover, the sex differences in evidence here are similar but not identical to cognitive sex differences observed in other mammalian species.

As in many studies of rodents and humans, male and female rhesus monkeys performed a task equally proficiently when multiple cues assisted their search. The elimination of colored markers that directly indicated goal location produced only slight impairments in initial performance that disappeared before four trials on the new task were completed. Contrary to predictions, the initial impairment was greater in male subjects than in females. Yet, within females, prenatal hormone manipulations were associated with effects in the hypothesized direction with exogenous androgen treatment late in gestation producing the best performance and flutamide treatment hindering performance. When instead, spatial information was made unreliable and only landmarks indicated food location, performance of both sexes dropped markedly. Again, a sex difference in performance was evident, but this effect was in the predicted direction, with females outperforming males. Performance of males treated with flutamide early in gestation was more like that of females and differed significantly from control males. Within females, late in gestation flutamide treatment consistently produced the best performance.

The lack of a consistent sex difference on Dual Cue Acquisition concurs with multiple studies in humans demonstrating no sex differences when cues are numerous (Leprow et al., 2000; Sandstrom et al., 1998). Studies of rodents, however, often find a male advantage even when both landmark and spatial information are available (Seymour et al., 1996, Williams et al., 1990). The cues available in the study discussed here differ from each of those referenced in that markers directly identified correct locations. In cited studies, the only landmarks available were distal to goal locations, providing cues for navigation but not immediate identification of correct locations. Nevertheless, the findings suggest the possibility that a male advantage in remembering food locations regardless of the presence of multiple cues may be present in rodents but not in primates, whether human or nonhuman. Although rodent research often informs primate research, it should not be surprising that there are species-specific sex differences in cognition. With their different social ecology, greater cognitive complexity, different scales of movement, and different preferred sensory modalities, primate cognitive sex differences are likely to differ, and perhaps be less generalized, than sex differences in rodents.

Performance of both males and females was scarcely affected by the removal of colored markers in the Spatial Task. Female scores across the ST actually exceeded the scores on the previous block of trials, likely due to continued improvement from Dual Cue Acquisition. However, a sex difference in performance was revealed, in the opposite direction to that expected. Females outscored males across the four trials, and females showed a smaller drop in scores from the previous trial to the initial ST trial than did males. No known previous rodent or human study has found a female advantage on a similar task.

However, important distinctions between the current study and previous work exist. In our Spatial Task, the most salient markers were removed, but numerous visual cues remained, including the researcher's vestibule, the capture area, and environmental cues outside of the testing area. When testing humans on a virtual water maze, men outperformed women even in the presence of distal landmarks in most cases (Astur et al., 1998; Driscoll et al., 2005; Sava et al., 2002), but in one case, a sex difference was only revealed when such landmarks were removed (Sandstrom et al., 1998). When all landmarks were removed but the geometry of the testing area remained constant on a radial arm maze, both male and female rats were unaffected (Williams et al., 1990), much like the female, but not male, monkeys. Thus, the impairment observed in our male subjects when markers were removed is unique, but so was the testing situation. Analysis of Marker Task performance indicates that neither males nor females relied on the presence of the colored markers, and one possibility is that the initial drop in male scores on the ST is more an impairment due to an observed change in the environment and not a specific lack of knowledge about the correct locations.

The female performance advantage was clearest on the Marker Task. Although both males and females did not readily visit the marked locations, female, but not male, visits to these correct locations exceeded chance. This sex difference was in the predicted direction and concurs with evidence in both rodents (Kanit et al., 1998; Williams et al., 1990) and humans (Galea and Kimura, 1993; Lawton, 1994, 1996; Saucier et al., 2002; Ward et al., 1986). Although females clearly learned the location of baited goal boxes during Dual Cue Acquisition, as evidenced by their excellent Spatial Task performance, they also apparently learned the relevance of the markers during acquisition. Because scores on the Marker Task were higher than scores on the initial Dual Cue Acquisition trials for females, learning during the Marker Task alone does not likely account for the better than chance performance.

The poorer than chance performance by the male subjects may be explained by a male tendency to perseverate. Across the four Marker Task trials, males continued to visit previously correct goal box locations at a steady rate, whereas visits to these same goal boxes decreased in females. According to at least one rodent study, females perform better than males on a reversal task, in which stimuli are paired with specific responses and then the pairings are reversed (Guillamón et al., 1986), suggesting that males are less behaviorally flexible (although the direction of this sex difference depends on task specifics (Wong and Judd, 1973)). Other studies, while not measuring perseveration *per se*, have found evidence of a greater inflexibility in male than female subjects (Kanit et al., 1998; Williams et al., 1990). Although subjects in this study showed no evidence of using motor strategies or biases to find goal locations (except perhaps for the first goal box visit), male rats are more likely to rely on motor strategies than are females, affecting their flexibility (Williams et al., 1990; Williams and Meck, 1991). Thus, both greater attention to landmarks by females and less flexibility in males may have contributed to the sex difference observed on the Marker Task in the current study, resulting in enhanced female performance.

Throughout testing, both males and females committed surprisingly few working memory errors. The only previously published study showing a sex difference in spatial cognition in rhesus monkeys utilized a computerized test of spatial working memory (Lacreuse et al., 1999), and a recent study of squirrel monkey cognition using a very similar facility reported numerous working memory errors (Ludvig et al., 2003). However, the procedures utilized by the latter study differed markedly from those presented here. In the squirrel monkey study, only a single food pellet was placed in each correct goal box. Thus, the working memory errors reported by his group are return visits to correct goal boxes. The parallel to the working memory errors reported in our study, return visits within a trial to incorrect locations, are not reported in Ludvig's study. Also, the testing procedure of Ludvig and colleagues required ten trials within a single day's session. Five seconds after a subject had retrieved all food items from

goal locations, a new trial began with the correct locations automatically rebaited while the subject remained in the facility. This protocol very likely increased the likelihood of working memory errors by blurring the distinction between individual trials. In our study, animals received only one trial per day, thereby creating distinct divisions between individual trials.

The ease with which the subjects performed the Spatial Task compared to the difficulty that the Marker Task presented was surprising. It is easily argued that the Spatial Task was easier than the typical “geometric” test in which spatial information is limited to room shape and does not include distal landmarks or a stable start location. However, the markers used in the Marker Task were also unique in that they were directly associated with correct locations; in fact, the salience of these blue disks initially caused concern that the Marker Task would prove too easy for subjects. Yet, this is not the first study to demonstrate a strong tendency for animals to preferentially use geometric cues and fail to use landmark cues. Wild capuchins easily learned the spatial locations of platforms holding bananas but had difficulty when yellow markers but not consistent location identified the placement of the bananas (Garber and Paciulli, 1997). Male rats preferentially use place cues and largely ignore markers in finding food rewards even when markers are equally reliable and informative (Gibson and Shettleworth, 2003). In the natural environment, spatial or geometric cues may be more consistent and reliable than the presence of a single given landmark, resulting in the evolution of a preference to use spatial cues instead of landmarks.

The effects of the prenatal androgen manipulations in this study were limited, but nevertheless, informative. Exogenous androgen prenatally administered to females did not consistently masculinize their performance. In fact, if anything, prenatal androgen treatment to females enhanced their performance in a female direction. On both the Spatial Task and the Marker Task, females outperformed males, and androgen treatment late in gestation was associated with the best performance of all groups. Thus, prenatal androgen treatment of females was associated with the strongest performances in a female-like direction. Studies of morphological and behavioral endpoints in these females since birth have largely failed to detect effects of the exogenous prenatal androgen treatments (Herman et al., 2000; Herman et al., 2003; Zehr et al., 2005, but see Tomaszewski et al., 2001), and our treatments were associated with maternal androgen levels that were one-sixth that of treated mothers in previous studies, which produced obvious morphological and behavioral effects (Goy et al., 1988; Goy and Resko, 1972; Goy et al., 1989; Herman et al., 2000; Resko et al., 1987). Nevertheless, the treatment likely mimicked variations in androgen exposure experienced by fetal females due to endogenous or environmental factors. Our results therefore suggest that these relatively small perturbations in prenatal androgen exposure may produce enhancements of cognitive performance, without exceeding the range of normal females.

The extent to which fetal females are masculinized by their naturally low levels of androgen during gestation is unknown. Therefore, blocking endogenous prenatal androgen in females with flutamide treatment might either produce “super-females” or might have no effect. In no case in this study did flutamide-treated females differ significantly from control females, suggesting the latter possibility. However, late-gestation flutamide treatment was associated with the best Marker Task performance, providing some evidence of less defeminization in the LFFs. On the Spatial Task, however, LFFs showed some indication of drops in performance with the initial ST trial, similar to male subjects.

The Marker Task produced the clearest difference in performance between the sexes, with a definitive female advantage, and prenatal treatment of males with flutamide reduced this difference. Early in gestation flutamide treatment in males was associated with better MT performance as evidenced by significantly greater performance scores than control males. Also,

whereas control male scores were significantly worse than chance, EFM performance scores did not differ from chance.

Unfortunately, small sample size prevented analysis of the effects of late gestation flutamide treatment in males. Although five LFM's originally entered the study, three failed to complete training or testing. Of the two remaining, one performed well, but the other was a poor performer who spent his test periods pacing and circling within the facility. This association between flutamide treatment and difficulty in completing testing was greatest in the late-treated males but also extended to early-treated males and late-treated females. The problems with testing these subjects could be attributed to low motivation, high anxiety, or high distractability. Such difficulties were almost completely absent in animals that had not received flutamide treatment. The fact that difficulties in testing were observed in both sexes suggests that either 1) an effect of flutamide independent from its blockade of androgen receptors is responsible for the observed association or 2) prenatal androgen acts in both sexes to somehow increase behavioral traits that affect adaptations to the experimental testing conditions, a sort of "testability." This latter possibility would require that decreasing androgen exposure below the norm for either sex, and not just below a particular threshold level, results in greater anxiety or distraction or less motivation. The mechanism by which prenatal androgen perturbations might alter these factors is unknown.

In summary, prenatal androgen manipulations in female subjects produced only subtle alterations of cognitive performance, providing little support for the findings in humans (Berenbaum, 2001) and rodents (Isgor and Sengelaub, 1998; Joseph *et al.*, 1978; Roof, 1993; Roof and Havens, 1992; Williams *et al.*, 1990) that excess prenatal or neonatal sex steroids masculinize spatial abilities in females. However, the strong performance of all subjects on the Spatial Task renders interpretation of the negative finding problematic. In contrast, early-gestation flutamide treatment of male subjects clearly reduced reliance on local markers. This effect is as hypothesized and suggests that, as in rats (Williams *et al.*, 1990), early steroid exposure reduces the attention paid to visual markers in a search task.

Evidence collected in this study suggests that the sex difference in attention paid to and use of local markers in solving search tasks is present not only in rodents and humans, but also in rhesus monkeys. Sufficient evidence in rodents indicates that this tendency is eliminated by developmental androgen. The present results suggest that this is true for rhesus monkeys also, and may also be the case for their close relatives, humans. On the testing paradigm used here, female rhesus monkeys were unimpaired by the removal of local markers and instead performed slightly better than their male counterparts. This sex difference can largely be attributed to an initial drop in performance by male but not female subjects when colored markers were removed, and appears to be a unique example of greater spatial performance by adult females than by adult males. However, excellent performance by males on subsequent Spatial Task trials suggests that the drop in scores may not reflect a spatial ability impairment *per se*. Nevertheless, neither rhesus monkey life histories nor social organization predicts better female than male performance on such a task. Studies of cognitive sex differences in monkeys are too rare to establish whether this finding is unique to the testing paradigm or a characteristic of the species.

Acknowledgements

We thank Shannon Stephens, Lora R. Hunter, Jessica Raper, Stacy York, Jennie Crosby, Rhiannon Brey, and Franklynn Graves for their help with aspects of this study. Research was supported in part by NIH grants R01-MH50268, K02-MH01062 (K.W.), by NIH NRSA grant 1 F31 MH070143-01A1 (R.A.H.), by the Center for Behavioral Neuroscience under the STC Program of the National Science Foundation under Agreement No. IBN-9876754, and by NCRR grant RR-00165 to the Yerkes National Primate Research Center which is fully accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International.

References

- Astur RS, Ortiz ML, Sutherland RJ. A characterization of performance by men and women in a virtual Morris water task: a large and reliable sex difference. *Behav. Brain Res* 1998;93:185–190. [PubMed: 9659999]
- Balda RP, Kamil AC. The spatial memory of Clark's nutcrackers (*Nucifraga columbiana*) in an analogue of the radial arm maze. *Anim. Learn. Behav* 1988;16:116–122.
- Berenbaum SA. Cognitive function in congenital adrenal hyperplasia. *Endocrinol. Metabol. Clin. North Amer* 2001;30:173–192.
- Cramer AE, Gallistel CR. Vervet monkeys as traveling salesmen. *Nature* 1997;387:464–464. [PubMed: 9168107]
- Dell'Omo G, Ricceri L, Wolfer DP, Poletaeva II, Lipp HP. Temporal and spatial adaptation to food restriction in mice under naturalistic conditions. *Behav. Brain Res* 2000;115:1–8. [PubMed: 10996402]
- Driscoll I, Hamilton DA, Yeo RA, Brooks WM, Sutherland RJ. Virtual navigation in humans: the impact of age, sex, and hormones on place learning. *Horm. Behav* 2005;47:326–335. [PubMed: 15708762]
- Galea LAM, Kavaliers M, Ossenkopp KP, Innes D, Hargreaves EL. Sexually dimorphic spatial learning varies seasonally in 2 populations of Deer Mice. *Brain Res* 1994;635:18–26. [PubMed: 8173954]
- Galea LAM, Kimura D. Sex differences in route-learning. *Pers. Individ. Differ* 1993;14:53–65.
- Garber PA, Paciulli LM. Experimental field study of spatial memory and learning in wild capuchin monkeys (*Cebus capucinus*). *Folia Primatol* 1997;68:236–253. [PubMed: 9360308]
- Gibson BM, Shettleworth SJ. Competition among spatial cues in a naturalistic food-carrying task. *Learning and Behavior* 2003;31:143–159. [PubMed: 12882373]
- Goy RW, Bercovitch FB, McBair MC. Behavioral masculinization is independent of genital masculinization in prenatally androgenized female rhesus macaques. *Horm. Behav* 1988;22:552–571. [PubMed: 3235069]
- Goy RW, Resko JA. Gonadal hormones and behavior of normal and pseudohermaphroditic nonhuman female primates. *Recent Prog. Horm. Res* 1972;28:707–733. [PubMed: 4631621]
- Goy, RW.; Uno, H.; Sholl, SA. Psychological and anatomical consequences of prenatal exposure to androgens in female rhesus. In: Mori, T.; Nagasawa, H., editors. *Toxicity of Hormones in Perinatal Life*. CRC Press, Inc.; Boca Raton, Florida: 1989. p. 127-142.
- Guillamon A, Valencia A, Cales JM, Segovia S. Effects of early postnatal gonadal steroids on the successive conditional discrimination reversal learning in the rat. *Physiol. Behav* 1986;38:845–9. [PubMed: 3823204]
- Herman RA, Jones B, Mann DR, Wallen K. Timing of prenatal androgen exposure: Anatomical and endocrine effects on juvenile male and female rhesus monkeys. *Horm. Behav* 2000;38:52–66. [PubMed: 10924287]
- Herman RA, Measday MA, Wallen K. Sex differences in interest in infants in juvenile rhesus monkeys: Relationship to prenatal androgen. *Horm. Behav* 2003;43:573–583. [PubMed: 12799175]
- Herman RA, Zehr JL, Wallen K. Prenatal androgen blockade accelerates pubertal development in male rhesus monkeys. *Psychoneuroendocrinology* 2006;31:118–130. [PubMed: 16112492]
- Holding CS, Holding DH. Acquisition of route network knowledge by males and females. *J. Gen. Psychol* 1989;116:29–41.
- Hubener F, Laska M. A two-choice discrimination method to assess olfactory performance in pigtailed macaques, *Macaca nemestrina*. *Physiol. Behav* 2001;72:511–9. [PubMed: 11282134]
- Isgor C, Sengelaub DR. Prenatal gonadal steroids affect adult spatial behavior, CA1 and CA3 pyramidal cell morphology in rats. *Horm. Behav* 1998;34:183–198. [PubMed: 9799628]
- Kamil AC, Balda RP, Olson DJ. Performance of four seed-caching Corvid species in the radial arm maze analog. *J. Comp. Psychol* 1994;108:385–393. [PubMed: 7813195]
- Kanit L, Taskiran D, Furedy JJ, Kulali B, McDonald R, Pogun S. Nicotine interacts with sex in affecting rat choice between “look-out” and “navigational” cognitive styles in the Morris water maze place learning task. *Brain Res. Bull* 1998;46:441–445. [PubMed: 9739007]

- Khan Z, Herman RA, Wallen K, Balch T. An outdoor 3-d visual tracking system for the study of spatial navigation and memory in rhesus monkeys. *Behav. Res. Methods* 2005;37:453–463. [PubMed: 16405140]
- Lacreuse A, Herndon JG, Killiany RJ, Rosene DL, Moss MB. Spatial cognition in rhesus monkeys: Male superiority declines with age. *Horm. Behav* 1999;36:70–76. [PubMed: 10433888]
- Lawton CA. Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles* 1994;30:765–779.
- Lawton CA. Strategies for indoor wayfinding: The role of orientation. *J. Environ. Psychol* 1996;16:137–145.
- Leplow B, Holl D, Zeng LJ, Mehdorn M. Investigation of age and sex effects in spatial cognitions as assessed in a locomotor maze and in a 2-D computer maze. *Spatial Cognition II* 2000;1849:399–418.
- Linn MC, Petersen AC. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Dev* 1985;56:1479–1498. [PubMed: 4075870]
- Ludvig N, Tang HM, Eichenbaum H, Gohil BC. Spatial memory performance of freely-moving squirrel monkeys. *Behav. Brain Res* 2003;140:175–83. [PubMed: 12644290]
- McDowell AA, Brown WL, McTee AC. Sex as a factor in spatial delayed-response performance by rhesus monkeys. *J. Comp. Physiol. Psychol* 1960;53:429–432.
- Menzel EW. Chimpanzee spatial memory organization. *Science* 1973;182:943–945. [PubMed: 17737534]
- Newcombe N, Bandura MM, Taylor DG. Sex difference in spatial ability and spatial activities. *Sex Roles* 1983;9:377–386.
- Olson DJ, Kamil AC, Balda RP. Effects of response strategy and retention interval on performance of Clark nutcrackers in a radial maze analog. *J. Exp. Psychol.-Anim. Behav. Process* 1993;19:138–148. [PubMed: 8505594]
- Resko JA, Buhl AE, Phoenix CH. Treatment of pregnant rhesus macaques with testosterone propionate: observations on its fate in the fetus. *Biol. Reprod* 1987;37:1185–1191. [PubMed: 3442696]
- Resnick SM, Gottesman II, Berenbaum SA, Bouchard TJ. Early hormonal influences on cognitive functioning in Congenital Adrenal Hyperplasia. *Dev. Psychol* 1986;22:191–198.
- Roof RL. Neonatal exogenous testosterone modifies sex difference in radial arm and Morris water maze performance in prepubescent and adult rats. *Behav. Brain Res* 1993;53:1–10. [PubMed: 8466654]
- Roof RL, Havens MD. Testosterone improves maze performance and induces development of a male hippocampus in females. *Brain Res* 1992;572:310–313. [PubMed: 1611529]
- Sandstrom NJ, Kaufman J, Huettel SA. Males and females use different distal cues in a virtual environment navigation task. *Cognit. Brain Res* 1998;6:351–360.
- Saucier DM, Green SM, Leason J, MacFadden A, Bell S, Elias LJ. Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behav. Neurosci* 2002;116:403–10. [PubMed: 12049321]
- Sava S, Tropp J, Astur RS, Constable RT, Markus EJ. Sex differences and behavioral correlates in mental rotation and virtual maze tasks [abstract]. *Horm. Behav* 2002;41:488.
- Seymour P, Dou H, Juraska JM. Sex differences in radial maze performance: Influence of rearing environment and room cues. *Psychobiology* 1996;24:33–37.
- Tomaszycki ML, Davis JE, Gouzoules H, Wallen K. Sex differences in infant rhesus macaque separation-rejection vocalizations and effects of prenatal androgens. *Horm. Behav* 2001;39:267–276. [PubMed: 11374912]
- Wallen K. Hormonal influences on sexually differentiated behavior in nonhuman primates. *Front. Neuroendocrinol* 2005;26:7–26. [PubMed: 15862182]
- Ward SL, Newcombe N, Overton WF. Turn left at the church, or 3 miles north: A study of direction giving and sex differences. *Environ. Behav* 1986;18:192–213.
- Williams CL, Barnett AM, Meck WH. Organizational effects of early gonadal secretions on sexual differentiation in spatial memory. *Behav. Neurosci* 1990;104:84–97. [PubMed: 2317288]
- Williams CL, Meck WH. The organizational effects of gonadal steroids on sexually dimorphic spatial ability. *Psychoneuroendocrinology* 1991;16:155–176. [PubMed: 1961837]

- Wong R, Judd M. Infantile handling and successive spatial reversal learning in rats. *Behav. Biol* 1973;8:391–7. [PubMed: 4705981]
- Zehr JL, Tannenbaum PL, Jones B, Wallen K. Peak occurrence of female sexual initiation predicts day of conception in rhesus monkeys (*Macaca mulatta*). *Reprod. Fertil. Dev* 2000;12:397–404. [PubMed: 11545179]
- Zehr JL, Van Meter PE, Wallen K. Factors regulating the timing of puberty onset in female rhesus monkeys (*Macaca mulatta*): Role of prenatal androgens, social rank, and adolescent body weight. *Biol. Reprod* 2005;72:1087–1094. [PubMed: 15625235]

Fig. 1.

Performance scores for rhesus monkey males, all females, and females without interfering infants across different conditions of a search task.

Fig. 2.

Change in performance scores from the previous trial to first Spatial Task trial in young adult rhesus monkeys who received various prenatal androgen manipulations. Abbreviations as in Table 1.

Fig. 3. Average performance scores on the four Marker Task trials in young adult rhesus monkeys who received various prenatal androgen manipulations. Abbreviations as in Table 1.

Fig. 4. Perseverative scores. Average perseverative performance scores for male and female rhesus monkeys on the Marker Task. Scores were calculated as if the original Dual Cue Acquisition goal boxes remained correct

Table 1
Treatment nomenclature and numbers of rhesus monkey subjects.

Treatment & Timing	Early Androgen		Late Androgen		Early Flutamide		Late Flutamide		Control							
	Female	EAF	Female	LAF	Male	EFM	Female	EFF	Male	LFM	Female	LFF	Male	CM	Female	CF
Total Attempted	5		6		7		6		5		7		9 ^a		6 ^b	
Total Included	5		6		5		6		2		6		9 ^a		6 ^b	

^aTwo of these males did not receive vehicle injections prenatally.

^bThree of these females did not receive vehicle injections prenatally.

Table 2 Prenatal treatments of rhesus monkey subjects failing to complete testing successfully or having numerous trials with no goal box visits. Abbreviations as in Table 1.

Abbreviation	EAF	LAF	EFM	EFF	LFM	LFF	CM	CF
Failure to habituate ^a			Bh6 Js5		Nv5	Cb6 Se6		
Failure to train			Bh6		Cd6 Ks5	Se6		
Failure to complete testing successfully			Js5		Nv5			
>5 trials with zero visits		Kc6	Id6		Ta6	Cb6 Zp5		

^aThese subjects were allowed to continue on with training and, if successfully trained, testing.