PRIMATES, 21(4): 498–509, October 1980

Individual Differences in Young Rhesus Monkeys: Consistency and Change

J. STEVENSON-HINDE, R. STILLWELL-BARNES, and M. ZUNZ
M. R. C. Unit on the Development and Integration of Behaviour

ABSTRACT. At 1 year and 2.5 years of age, rhesus monkeys were removed from the colony to a strange situation. Upon introduction, each monkey was watched for 2 hr, given a series of behavioural tests, and watched again six days after introduction. Activity increased with age, as did readiness to approach a ball. Time spent looking at a mirror decreased with age. Although test measures were reliable at each age, there was no significantly positive correlation from one age to the next. Nevertheless, with males, but not females, significant correlations occurred at each age between behaviour in the strange situation and observers' ratings of behaviour in the colony.

INTRODUCTION

Anyone who works with primates, or for that matter many other species, becomes conscious of individual differences. Too often such differences are regarded as noise, obscuring the issue that is the focus of attention. Yet individual characteristics and relationships may play a part in determining such outcomes as individual breeding success or differences in group structure. While there is anecdotal evidence for such effects (e.g., LAWICK-GOODALL, 1971; YAMADA, 1971; HARCOURT, 1978), it is also necessary to come to terms with the individual differences in a quantitative way. Three routes are open. One is to examine individual differences on the basis of observational measures (e.g., HINDE & SPENCER-BOOTH, 1971). Another is to quantify the impressions of observers who are familiar with the animals (e.g., BUIRSKI, PLUTCHIK, & KELLERMAN, 1978; STEVENSON-HINDE, STILLWELL-BARNES, & ZUNZ, 1980). The third is to assess individuals over a series of behavioural tests (e.g., SPENCER-BOOTH & HINDE, 1969; HINDE, LEIGHTON-SHAPIRO, & MCGINNIS, 1978).

As with the study of individual differences in humans, two related issues arise: (1) How consistent are the characteristics, over time and across situations?; and (2) At what level of analysis should consistencies be sought? (see e.g., BEM & ALLEN, 1974; BLOCK, 1977). By exploring the above three means of quantifying individual differences, this paper provides a first step towards answering these questions with non-human primates.

METHODS

SUBJECTS

The subjects were 1-year-old and 2.5-year-old rhesus monkeys (*Macaca mulatta*). They were born and reared in any one of six groups, each containing one adult male, several adult females, and their offspring. Each group lived in an outdoor pen ($5.5 \times 2.4 \times 2.5$ m) with access to an indoor room ($1.9 \times 1.3 \times 2.4$ m). The groups had only minimal disturbance, of a sliding door between the inside room and the outdoor pen being closed for feeding, cleaning, or behavioural observations.

Normally, infants received no handling until aged 1 year, when they were removed with mothers for a week of behavioural testing. The 25 1-year-olds fell into two groups:

Normals: These were ten females and nine males, who had never been separated from either their mothers or their group.

Miscellaneous: These were two females and four males, with various kinds of adverse early experiences, ranging from being orphaned to being removed from the colony.

At 2.5 years, infants were removed for testing on their own, without mothers. Four had not been tested at 1 year since their mothers were either about to give birth or had just given birth, and so could not have been removed from the colony; eleven others had been tested similarly at 1 year, but in a different building with a different recording system (HINDE, LEIGHTON-SHAPIRO, & MCGINNIS, 1978), and the remaining 16 had been tested in the present situation. The 31 2.5-year-olds fell into four groups:
Normals: These were nine females and five males, who had never been separated from mother.

Mother and infant removed and kept together: These were one female and five males, who had been removed from the colony with mother at age 30 weeks and kept with her for two weeks before being returned to the colony.

Mother and infant removed and kept apart: These were four females and three males, who had been removed from the colony with mother at age 32 weeks, put into separate buildings, and kept apart for two weeks before being returned to the colony (see HINDE, LEIGHTON-SHAPIRO, & MCGINNIS, 1978).

Miscellaneous: These were one female and three males, with various adverse early experiences, ranging from being orphaned to being removed from the colony.

PROCEDURE

At 1 year of age each infant and his mother, and at 2.5 years each infant on his own, was removed from the colony to a building which was used only for testing. It contained no other animals, although noises from the colony could be heard. The infant (or mother/infant pair) was placed in a wire-mesh cage ($1.0 \times 1.1 \times 1.1$ m), called the home cage, in which food pellets and water were available at all times. One of the wire-mesh walls was pulled back sufficiently to allow the infant but not mother access to an identical cage, called the test cage. The subjects were watched from another room through a one-way mirror, permanently mounted in a wall 1 m from the cage. Behaviour was recorded from either a keyboard or from solid state operant equipment into a WRATS computer-compatible system (WHITE, 1971).

Immediately upon introduction to the home cage, at 0930 hours on Day 0, behaviour was merely recorded for 2 hr, with no test objects present (Introductory Hut Watch). A similar watch, called the Last Hut Watch, occurred on Day 6. The tests themselves involved presentation of novel objects, the appearance of a masked person, or operant procedures. Except for the tests involving a masked person, it was necessary for the observer to enter the testing room before testing began, either to introduce an object into the test cage or to attach equipment. Objects were presented from the side of the test cage which was farthest from the home cage, and this will be referred to as the "far side" of the test cage. The observer was in the testing room for as short a time as possible (at most 2 min), and never looked directly at the monkeys.

The tests were as follows, with an asterisk indicating those which occurred only at age 2.5 years:

Ball test: A tan-coloured plastic ball (16 cm diameter) was placed in the centre of the test cage for 15 min.

Food test: Either three slices of banana, about 1 cm thick, were placed in a row on the floor of the test cage against its far side, or one date was placed in the centre of this far side. Behaviour was recorded for 15 min.

*Mask test**: A man was covered from neck to floor by a black cloth with white dots on it, and he wore a chimp mask. He entered the testing room, stood still 100 cm from the far side of the test cage for 1 min, and then left. Behaviour was recorded for 15 min from the time when the man stood still.

Mirror test: A mirror (46 cm square) was attached to bolts on the outside of the far side of the test cage. It was left there 15 min.

*Slide test**: A translucent screen (60 × 90 cm) with an operant lever (5 cm wide × 1 cm thick and protruding 2.5 cm) at its lower centre and 28 cm from the test cage floor was attached to bolts on the outside of the far side of the test cage. Each lever press operated a shutter in a slide projector which back-projected an image (56 × 35 cm) onto the screen. The shutter remained open for as long as the lever was held down. The first slide was available (by operating the lever which operated the shutter) for 60 sec from the first lever press. Lever pressing after that time produced a second slide, which in turn remained available for 60 sec, and so on.

There were a total of 16 different slides in all, of four poses of four different individuals: a 10-week-old male infant rhesus, an adult female rhesus, an adult male rhesus, and an adult brown Soay sheep. The slides were arranged in sets of four with each individual appearing once and in a different position within each set. The operant test ended either 60 sec after the 16th slide was earned or after 1 hr, whichever was the shorter. Immediately after this, any unearned slides were presented for 1 min each, with 1 min between slides.

On the following day, the Slide test was continued, using the same procedure as above, except that the slides were shown in the reverse order.

*Smartie test—reinforcement**: A stainless steel operant panel (60 × 90 cm) was attached to the bolts on the outside of the far wall of the test cage. It had a lever identical to that used in the Slide test at its lower centre, 42 cm from the test cage floor. Centred 6 cm above the lever was an opening through which a Smartie (similar to an M & M) could be delivered from a solenoid-operated dispenser at the back of the panel.

Three free Smarties, CRF, FR5—Since the aim of the training was not to shape the infants as quickly as possible, but to assess individual differences, a set procedure was followed: A free Smartie was delivered at the end of minutes 1, 2, and 3. Either at the end of this time or as soon as a lever press occurred (whichever happened first), each monkey was allowed to earn ten Smarties on a continuous reinforcement (CRF) schedule, in which each lever press produced one Smartie. After earning ten Smarties on CRF, the infant was allowed to earn ten Smarties on a fixed-ratio 5 schedule (FR5), in which every fifth press produced one Smartie. The operant panel was removed either upon completion of this or after 2 hr, whichever was shorter.

FR10—On the day after completion of FR5, the infant was allowed to earn ten Smarties on a fixed ratio 10 (FR10) schedule, in which every tenth press produced a Smartie. If all ten Smarties had not been earned in 2 hr, the procedure was continued on following days.

In all, a maximum of 8 hr, spread over four days, was allowed for completion of the entire task of earning 30 Smarties. If during CRF, FR5, or FR10, 30 min elapsed with no lever press, then up to three Smarties were delivered, each contingent upon orientation to the lever. These were not counted as part of the 30 reinforcements.

*Smartie test—extinction**: As soon as an infant had earned ten Smarties on FR10, 10 min of extinction followed. During this time, no Smarties were delivered at all, even though lever pressing occurred. At the end of 10 min, the operant panel was removed.

Calling the day of removal from the colony Day 0, the tests for 1-year-olds were scheduled

as follows: Ball test at 1200 hours on Days 3 and 5; Mirror test at 1200 hours on Days 4 and 6; and Food (banana) test at 1230 hours on Days 4 and 6.

The schedule for 2.5-year-olds was: on Days 3 and 5, Food (banana on Day 3 and date on Day 5) test at 1000 hours, Mirror test at 1200 hours, and Ball test at 1230 hours. The Mask test occurred at 1430 hours on Days 4 and 6. Smartie testing began at 1000 hours on Day 4 and continued on Days 7–9 until completed. Finally, the Slide test occurred at 1000 hours on Days 10 and 11.

RESULTS

One-Year-Olds

Test/Retest Consistency and Change

Table 1 indicates Spearman correlation coefficients and significant differences for tests and retests, with two measures from the Introductory Hut Watch (IHW) and Last Hut Watch (LHW) included as well. Activity in the watches, in which no test objects were present, was measured in terms of the frequency of "boxes" entered. Each cage was divided into eight imaginary cubes or "boxes," four upper and four lower. The frequency of boxes entered was divided by time off mother for the 1-year-olds, and by real time for the 2.5-year-olds who were in the hut alone. Activity while off mother was significantly lower in the IHW than in the LHW ($p<.001$, Wilcoxon matched-pairs test, two-tailed), with a median of 6 boxes per 100 sec off mother in the IHW, rising to 12 boxes per 100 sec off mother in the LHW. On the other hand, the frequency of whoo calls, usually associated with some distress (Hinde, 1977), was low and did not differ significantly between the IHW and LHW. The median was 1.4 every 15 min in the IHW and .5 in the LHW. In general, the strange situation initially produced a depression in behaviour, coupled with more time spent on mother (Stevenson-Hinde, Zunz, & Stillwell-Barnes, 1980). In addition, Spearman correlation coefficients between the IHW and LHW were not high: .30 for activity while off mother and .38 for whoo calls.

For the tests themselves, although 48 hr and other tests elapsed between test and retest, Spearman correlation coefficients were highly significant, and ranged from .50 for time spent lipsmacking to .79 for time spent looking at the mirror while in the test cage. In addition, these two measures did not change significantly from Mirror 1 to Mirror 2. Indeed, monkeys spent a considerable proportion of each 15-min test in the test cage looking at the mirror: a median of 13.6% in Mirror 1 and 14.5% in Mirror 2. They responded by lipsmacking, something they did not do at all in the Food or Ball tests. For the Ball tests, the test/retest correlation between latencies to touch the ball was meaningless, since 23 infants in Ball 1 and 18 in Ball 2 did not touch the ball at all. However, time spent in the test cage did increase from 3.0% in Ball 1 to 8.2% in Ball 2. The other significant change was that latencies to touch the banana decreased from a median of 5.1 to 0.0 sec. Indeed, by the second Food test, 17 infants had touched the banana within the time it took the observer to leave the test room and sit at the keyboard. As this implies, latencies to touch were lower in Food 1 than in Ball 1, and in Food 2 v. Ball 2 (Wilcoxon matched-pairs, $p<.001$, two-tailed).

Thus, for all three tests, behaviour was consistent from test to retest. Only two significant changes occurred. Latency to touch a slice of banana decreased to a median of zero in Food 2, and time spent in the test cage increased from Ball 1 to Ball 2. The results are compatible with the view that the infants were less frightened of the food than the ball, and in each case less

frightened by the retest. For the Mirror tests, looking and lipsmacking showed no sign of habituation from test to retest.

Males v. Females

When the six correlation coefficients in Table 1, column 1 were considered for the 13 males and 12 females separately, all 12 coefficients were positive. For five of the six measures, coefficients were higher for males than for females, but this was not significant (Sign test, two-tailed).

For all seven measures in Table 1, males were compared with females in both test and retest, making a total of 14 comparisons (Mann-Whitney U tests, two-tailed). Only one was significant: during Food test 1, the males had a median latency of 0 sec, while the females had a median latency of 8.2 sec ($p<.02$, two-tailed). Thus, the sexes did not differ markedly in terms of either consistency from test to retest or absolute levels of responding.

Early Experience

Nineteen of the 25 1-year-olds were raised with their mothers in the colony, and experienced no separation from their mothers or their groups. However, the remaining six did have some adverse early experience, such as being orphaned or removed from the colony.

For all seven measures in Table 1, the adverse group was compared with the controls, in both test and retest, making a total of 14 comparisons (Mann-Whitney U tests, two-tailed). None were significant. When the six test/retest correlations were considered separately for each group, all were positive. Furthermore, there was no tendency for one group to be more

Table 1. Correlations and changes at 1 year, from 1 to 2.5 years, and at 2.5 years.

	1 year (n = 25)	to (n = 16)	2.5 years (n = 31)
Activity	I H W .30 (⌒) L H W	.41 (ns) .36 (〈)	I H W .22 (⌒) L H W
Whoo calls	I H W .38 (ns) L H W	×(〉) −.19 (ns)	I H W × (⌒) L H W
Latency to touch	Food 1 .67* (⌣) Food 2	.47 (〈) × (ns)	Food 1 × (⌣) Food 2
Latency to touch	Ball 1 × (ns) Ball 2	× (〉) .42 (〉)	Ball 1 .75*(⌣) Ball 2
Time in test cage	Ball 1 .61* (⌒) Ball 2	.23 (〈) .14 (〈)	Ball 1 .59* (ns) Ball 2
Time looking while in test cage	Mirror 1 .79* (ns) Mirror 2	−.66* (〉) −.11 (〉)	Mirror 1 .66* (ns) Mirror 2
Time lipsmacking	Mirror 1 .50* (ns) Mirror 2	.14 (ns) −.06 (ns)	Mirror 1 .58* (⌣) Mirror 2

Spearman correlation coefficients: * indicates significant ($p<.05$, two-tailed); × unobtainable due to ties. Wilcoxon matched-pairs tests: (〈) indicates a significant difference ($p<.05$, two-tailed); (ns) not significant.

Table 2. The significant correlations between CONFIDENT, EXCITABLE, and SOCIABLE colony scores and six hut measures for 11 1-year-old males.

Colony score:	CONFIDENT	EXCITABLE		SOCIABLE
	.83**	.75*	−.80*	−.73*
Hut measure:	Food 1 latency to touch	Ball 2 time in test cage	LHW whoo calls	LHW activity

*$p<.05$; ** $p<.01$, two-tailed. (For 11 females, there were no significant correlations.)

consistent than the other, since three measures produced higher coefficients for one group, and three for the other.

Thus, over all the measures, no differences emerged between the group with unusual early experiences and the controls, as measured by absolute values or consistency from test to retest. However, this is not to say that adverse experience had no effect, for within this heterogeneous group of six, five individuals showed extreme behaviour, in one direction or the other. (STEVENSON-HINDE, ZUNZ, & STILLWELL-BARNES, 1980, Table 4).

Correlations with Colony Scores

Every November, two observers independently rated all monkeys in the colony, using a list of behaviourally-defined adjectives. For each year, their ratings were summed and converted to standardized scores, which were then combined to provide the following summary scores for each individual: CONFIDENT, EXCITABLE, and SOCIABLE (see STEVENSON-HINDE, STILLWELL-BARNES, & ZUNZ, 1980). Scores from that November after testing were available for 22 infants (one female was too ill, a male's mother had died between testing and rating, and another male and his mother had been removed from the colony). To reduce the number of measures to be correlated with the scores, only one test and not its repetition was used, to provide the following six measures (see Table 1): Of the IHW and LHW, the more settled LHW was chosen, to provide activity while off mother and frequency of whoos; from Food 1, the latency to touch; and from Mirror 1 duration of looking while in the test cage and duration of lipsmacking. Finally, since 11 infants did not enter the test cage during Ball 1, Ball 2 was chosen, to provide duration in test cage (latency to touch was omitted, since even in Ball 2, 18 infants did not touch the ball).

When these six measures were correlated with the three colony scores, 18 correlation coefficients resulted for the 11 males, and another 18 for the 11 females. For 12 of the 18 correlations, the absolute value of the coefficient was greater for males than for females. Furthermore, none of the females' coefficients were significant ($p<.05$, two-tailed), but four of the males' 18 coefficients were, and these are indicated in Table 2. It shows that males who were CONFIDENT in the colony had high latencies to touch the banana, with a correlation of .83. EXCITABLE males spent a long time in the test cage when the ball was there (.75), and during the LHM they whooed little (−.80). Finally, during the LHW, SOCIABLE males were not active while off mother (−.73).

TWO-AND-A-HALF-YEAR-OLDS

Test/Retest Consistency and Change

As at 1 year when mothers were present, at 2.5 years when mothers were absent, activity

increased from the IHW (a median of 6 boxes/100s) to the LHW (22 boxes/100s) ($p<.001$, Wilcoxon matched-pairs test, two-tailed). Furthermore, consistency from IHW to LHW was similar, with a Spearman correlation coefficient of .22, compared with one of .30 at 1 year. However, at 2.5 years, 28 out of 31 infants gave no whoo calls in the IHW, and whoo calling increased from this to the LHW ($p<.02$, two-tailed). Nevertheless, it was still low, with ten infants emitting no whoo calls in the LHW.

Test/retest coefficients were again highly significant ($p<.002$, two-tailed) and not different between the two age groups. Although time spent looking at the mirror while in the test cage did not change significantly from Mirror 1 (a median of 9.0%) to Mirror 2 (5.6%), time spent lipsmacking to the mirror decreased from the first to second test, as did latency to touch food (to 0 sec for 25 individuals) and latency to touch the ball (to a median of 54 sec.). Although the latencies of the 2.5-year-olds to touch the banana in Food 1, which was the very first test, were not significantly lower than those to touch the ball in Ball 1, latencies in Food 2 were lower than those in Ball 2 ($p<.001$, two-tailed), again suggesting that the ball was more frightening than food.

As stated above, activity during the LHW was higher than during the IHW. Similarly, activity during either Mask test (medians of 17 and 26 boxes/100s) was higher than activity during the IHW (Wilcoxon matched-pairs tests, $p<.001$, two-tailed), and no significant differences emerged between either Mask test and the LHW. This suggests that activity during the Mask test was neither depressed as in the IHW nor elevated above that in the LHW. However, 27 of the 31 monkeys did threaten the masked person, and time spent threatening did not decrease from Mask 1 to Mask 2.

The Operant Tests

The median time taken to earn ten Smarties on CRF, ten on FR5, and ten on FR10 was 158 min. Since each session lasted a maximum of 120 min, this implies that the reinforcement task was normally completed within the minimum of two sessions (i.e., one for CRF and FR5 and another for FR10). However, the range was great. One male, *CT*, completed the task in a total of 26 min, contrasted with five males and four females which did not complete the task within four 2-hr sessions. Their failure is not surprising since (1) food pellets and water were available at all times and (2) shaping was minimal, in order to maximize individual differences.

Thus, only 22 infants went on to the extinction procedure, which came into effect as soon as the FR10 procedure was completed. For these infants, the median total number of lever presses during the 10 min of extinction was 44, with a range from 15 to 322. The male, *CT*, who got through reinforcement most readily also stopped readily, giving only 26 presses in extinction. Nevertheless for all 22 infants, the correlation was negative (−.20) between time taken to earn 30 Smarties and the number of responses given in extinction.

Following extinction, at least one day elapsed before the Slide test began, and since most individuals finished extinction on Day 7, it was normally three days. Over the two daily sessions of the Slide test, the median total frequency of lever presses was 3, with a range from 0 to 229. The median total time the lever was held down was 5.0 sec, with a range from 0 to 639 sec. The median number of slides earned was 3, with a range from 0 to the maximum possible number of 32. Of the five infants who did not lever press at all, four were males. Of the five who earned all 32 slides, 16 on each day, all were females. One of these was *Rachel*, who produced the maximum number of lever presses (229) and held the lever down for the

longest time as well (639s). While holding the lever down with a foot, she explored the image with her hands and mouth. She had an above-average EXCITABLE score, as did all five females who earned 32 slides. Their scores ranged from .38 to 1.10, with *Rachel* having the median of the five at .63 (Since these are based on standardized scores, the colony mean was zero). *Rachel* had two contemporaries in her pen, and her group was peculiar in that these three were the only non-adults in it. The female contemporary, *Phoebe*, also earned all 32 slides, pressing the lever 151 times and holding it down for 421s. The other, *Gareth*, was the best of the males on the Slide test, earning 27 slides. Thus, the Slide test produced a wide range of operant performance, which is possibly related to sex and the nature of the group from which an infant came.

Males v. Females

When the four measures during the watches and 12 measures during the tests (selecting the first test only for Food, Mirror, and Mask and the second only for Ball, since 14 infants did not touch the ball in Ball 1 but only six did not in Ball 2) were compared for the 16 males and 15 females, no significant differences emerged (Mann-Whitney U tests, two-tailed). When the seven test/retest coefficients (see the five measures in Table 1, column 3 plus activity and threats during the Mask tests) were considered separately for males and females, all were positive, with no significant tendency for those of one sex to be higher than the other (i.e., 3/7 were higher for males than for females). Thus, over all the measures, no significant differences emerged between males and females. Of course, sex may be interacting with other variables, as suggested by the above results with the Slide test.

Before proceeding further, the number of measures had to be reduced. Firstly, the unsettled IHW was discarded, as well as two of the three Slide test measures, since the intercorrelations were .96, .97, and .99. The total duration of lever pressing was chosen, since it represented viewing time. For the Ball test, the latency to touch Ball 2 was selected, since only 14 monkeys touched the ball in Ball 1. Also, time spent in the test cage was discarded as being less relevant than actually touching the ball. Except for Ball 2, all other retests were omitted, leaving the following measures: activity (LHW); frequency of whoo calls (LHW); latency to touch (Food 1); latency to touch (Ball 2); time looking while in test cage (Mirror 1); time lipsmacking (Mirror 1); activity (Mask 1); time threatening (Mask 1); duration of test (Smartie Reinforcement); frequency lever press (Smartie Extinction); and duration lever press (Slides).

A principal component analysis of these 11 measures, each converted to ranks across the 31 infants, produced five components with eigenvalues greater than one. In addition to producing this many components, the groupings did not make sense, even with rotation. For example, latency to touch food, time spent lipsmacking to the mirror, and time spent holding the lever for slides all loaded highly (.70 or greater) on Component II. While activity during the Mask test loaded highly (.95) on Component I, activity in the LHW did not (—.16). However, the Spearman correlation was highly significant (.63). Thus, the above test measures were used on their own, with the omission of the LHW measures (activity correlated with Mask test activity and whoo calling was infrequent).

Adverse Early Experience

For the remaining nine measures, the three groups—controls ($n = 14$), mother and infant removed from the colony and kept together ($n = 6$), and mother and infant removed and kept apart ($n = 7$)—permitted 27 comparisons (i.e., Mann-Whitney U tests). However, only

Table 3. The significant correlations between CONFIDENT, EXCITABLE, and SOCIABLE colony scores and nine hut measures for 16 2.5-year-old males.

Colony score:	CONFIDENT		EXCITABLE		SOCIABLE
	−.50*	−.63*	−.54*	.50*	.51*
Hut measure:	Mirror 1 *d* look while in test cage	Mirror 1 *d* lipsmack	Smartie rft *d* of test	Slides *d* lever	Slides *d* lever

* $p<.05$, two-tailed; *d* is for duration. (For 15 females, there were two significant correlations, see text.)

one significant difference emerged: the controls spent less time looking at the mirror while in the test cage than did the mother and infant removed and kept apart group ($p<.05$, two-tailed). Of the six test/retest correlation coefficients for each group, there was no tendency for any group to be more or less consistent than any other.

This lack of difference between controls and mother and infant removed groups is consistent with earlier results (HINDE, LEIGHTON-SHAPIRO, & MCGINNIS, 1978). There it was concluded that an early separation experience can, but need not, produce long-term effects. The only type of separation experience found to give long-term sequelae involved removal of the mother from the colony, leaving the infant behind—a treatment not represented in the present sample.

Correlations with Colony Scores

The present tests at 2.5 years were carried out during the months of October to April. From the November of that period, scores based on observer ratings were available for all 31 infants. Thus, three colony scores, CONFIDENT, EXCITABLE, and SOCIABLE, were correlated with the nine test measures listed above. Over all the measures, males did not produce higher correlations than females: in only 13 out of 27 cases were the absolute values of the coefficient greater for males than for females. However, only two of the females' 27 Spearman correlation coefficients were significant: females who were CONFIDENT in the colony took a long time to complete the Smartie reinforcement test in the hut (.66; $p<.01$, two-tailed), as did females who were SOCIABLE (.52; $p<.05$, two-tailed). As this implies, there was an unusually high correlation between CONFIDENT and SOCIABLE scores of these females (.86), which did not hold for the males (.26).

Nevertheless, the males produced 5/27 significant correlations ($p<.05$, two-tailed), which are shown in Table 3. Males with high CONFIDENT scores in the colony were not responsive to the mirror: the correlation with time spent looking while in the test cage was −.50, and with time spent lipsmacking was −.63. On the other hand, EXCITABLE males did well in operant tests: the correlation with time to complete Smartie reinforcement was −.54, and the correlation with time spent holding down the lever in the Slide test was .50. Similarly, SOCIABLE males spent a long time holding down the lever for slides (.51).

CHANGE AND CONSISTENCY FROM 1 TO 2.5 YEARS

Sixteen individuals were tested in this situation at both ages. Long-term changes from 1 to 2.5 years may therefore be seen in the context of short-term change within each age as follows (Table 1): Activity increased both in the short term from the IHW to the LHW at each age, and in the long term from the 1-year LHW to the 2.5-year LHW. However, with the IHW, activity did not increase with age, suggesting that it was especially depressed at 2.5, when the

monkeys were introduced to the hut without their mothers. Whoo calling was also depressed during the 2.5-year IHW. That is, it was lower than during either the preceding 1-year IHW or the following 2.5-year LHW.

With the tests, latency to touch food decreased from Food 1 to Food 2 at each age. The increase from 1 year to 2.5 was probably not due to an age change, but to the fact that Food test 1 was the first test given at 2.5 years. With the Ball tests, time in the test cage increased from Ball 1 to Ball 2 at 1 year, as well as from 1 to 2.5 years. Similarly, latency to touch the ball decreased with age. It also decreased from Ball 1 to Ball 2 at 2.5 years.

However, with the mirror, less interest was shown with age, in that time spent looking at the mirror decreased from 1 to 2.5 years. It did not change within either age period. Finally, time spent lipsmacking to the mirror did not decrease within the 1-year test, and remained at the same level for the first 2.5-year test, but dropped significantly for the last 2.5-year test.

Thus, time spent looking at the mirror decreased only in the long term, while time spent lipsmacking decreased only in the short term (at 2.5 years). Both short-term and long-term changes in the same direction occurred as follows: activity increased, latency to touch the ball decreased, and time with the ball increased.

The correlations for each measure were higher in the short term than in the long term, as might be expected. One exception was activity, where they were as high over 1.5 years as over 6 days (i.e., from IHW to LHW). However, the low correlations within each age possibly reflect the different nature of the stressful IHW compared with the more settled LHW. The other exception was time spent looking at the mirror, where although the short-term correlations were significantly positive, three out of four long-term ones were negative, one being significantly so.

DISCUSSION

When tests were given to rhesus monkeys at 1 and at 2.5 years of age, behavioural measures were significantly reliable within each age but not across ages (Table 1). Latencies to touch indicated that a ball was more frightening than was food, and behaviour towards the ball increased with age. Conversely, time spent looking at a mirror decreased with age. In addition, this measure, which was significantly positively correlated from test to retest at each age, was significantly negatively correlated across ages, implying that those who looked at the mirror most at 1 year looked least at 2.5 years.

Of particular interest were the operant tests, which were given only at 2.5 years. They were correlated with scores from the colony (Table 3) and produced a wide range of performance. For example, while one male earned all 30 Smarties over a total period of only 26 min, five males and four females failed to complete the task within four 2-hr sessions. In extinction, the number of lever presses given in 10 min ranged from 15 to 322. During the Slide test, five individuals, four of them males, did not press for slides at all. At the other extreme, five infants, all of them females, earned the maximum of 32 slides. In spite of this, only one significant sex difference emerged on any of the tests at either age.

When scores based on observers' ratings in the colony (STEVENSON-HINDE, STILLWELL-BARNES, & ZUNZ, 1980) were correlated with behaviour in the hut, males showed more significant correlations than did females at each age. That is, between three colony scores and eight hut watch measures taken at 1 year, there were seven significant correlations out of 24 for males but none for females (STEVENSON-HINDE, ZUNZ, & STILLWELL-BARNES, 1980, Figure

III). Between colony scores and 1-year tests (excluding the watches), there were two significant correlations out of 12 for males but none for females (Table 2). Finally, between colony scores and 2.5-year tests, there were five significant correlations out of 27 for males compared with only two for females (Table 3).

The lack of cross-situational correlations for females is surprising since (1) each hut measure on its own was either reliable (i.e., the test measures) or was taken over a sufficiently long period to be representative (i.e., the hut watch measures), (2) the observer ratings were both reliable between observers and correlated meaningfully with earlier social behaviour (STEVENSON-HINDE, STILLWELL-BARNES, & ZUNZ, 1980), and (3) the use of colony scores based on observers' judgments, rather than isolated behavioural measures should maximize the possibility of finding significant cross-situational correlations (see e.g., BLOCK, 1977). However, before dismissing the females as unpredictable, another set of correlations raises the possibility that the two sexes may show cross-situational consistency in different ways. At 1 year, three discrete measures (time off mother, activity while off, and whoo calling) were taken over three 2-hr watches in the colony and the relatively settled last hut watch. Of the resulting correlations for each measure between the colony and the hut, none was significant for males, while whoo calling was highly significant for females ($p<.01$ two-tailed; STEVENSON-HINDE, ZUNZ, & STILLWELL-BARNES, 1980, Table II). This raises the possibility that the present lack of correlation between colony scores and hut behaviour for females reflects a failure to look at appropriate measures rather than a characteristic of the females themselves. For example, had similar tests been given in the hut and the colony, those colony measures might be correlated with hut measures for females but not males.

Nevertheless, the lack of any significant correlation between like measures for the males emphasizes the different nature of the two situations, one a group in which the monkey was raised and the other a strange situation without other monkeys (except mother at 1 year). Indeed, the cross-situational correlations that did emerge for males were not straightforward. For example, it was not the CONFIDENT but the EXCITABLE males who spent time in the test cage with the ball, whooed least, and did well on both the Smartie and Slide tests. Males who were CONFIDENT in the colony had a high latency to touch food and showed little interest in the mirror in the test situation (Tables 2 & 3). Thus the search for consistency may take unexpected paths, which in turn could be different for males than for females.

Acknowledgements. We are grateful to R. A. HINDE for his advice and helpful comments on the manuscript, to R. MCNAMEE for her statistical advice, to M. J. A. SIMPSON for catching the monkeys and recording some of the Mask tests, and to J. JOLLEY for catching the monkeys and looking after them so reliably well, as well as being the masked person in the Mask tests.

REFERENCES

BEM, D. J. & A. ALLEN, 1974. On predicting some of the people some of the time: The search for cross-situational consistencies in behaviour. *Psychol. Rev.*, 81: 506–520.

BLOCK, J., 1977. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In: *Personality at the Crossroads*, D. MAGNUSSON & N. S. ENDLER (eds.), Wiley, New York, pp. 37–63.

BUIRSKI, P., R. PLUTCHIK, & H. KELLERMAN, 1978. Sex differences, dominance, and personality in the chimpanzee. *Anim. Behav.*, 26: 123–129.

HARCOURT, S., 1978. Aspects of the behaviour of the mountain gorilla. Ph. D. thesis, Cambridge Univ., Cambridge.

HINDE, R. A., 1977. Mother-infant separation and the nature of inter-individual relationships: Experiments with rhesus monkeys. *Proceed. Royal Soc. London*, 196: 29–50.

————, M. E. LEIGHTON-SHAPIRO, & L. MCGINNIS, 1978. Effects of various types of separation experiences in rhesus monkeys five months later. *J. Child Psychol. & Psych.*, 19: 199–211.

———— & Y. SPENCER-BOOTH, 1971. Towards understanding individual differences in rhesus mother-infant interaction. *Anim. Behav.*, 19: 165–173.

LAWICK-GOODALL, J. VAN, 1971. *In the Shadow of Man.* Houghton Mifflin Co., Boston, Massachusetts.

SPENCER-BOOTH, Y. & R. A. HINDE, 1969. Tests of behavioural characteristics for rhesus monkeys. *Behaviour*, 33: 180–211.

STEVENSON-HINDE, J., R. STILLWELL-BARNES, & M. ZUNZ, 1980. Subjective assessment of rhesus monkeys over four successive years. *Primates*, 21: 66–82.

————, M. ZUNZ, & R. STILLWELL-BARNES, 1980. Behaviour of one-year old rhesus monkeys in a strange situation. *Anim. Behav.*, 28: 266–277.

WHITE, R. E. C., 1971. WRATS: A computer-compatible system for automatically recording and transcribing behavioural data. *Behaviour*, 40: 135–161.

YAMADA, M., 1971. Five natural troops of Japanese monkeys on Shodoshima Island. II. A comparison of social structure. *Primates*, 12: 125–150.

— Received *July 23, 1979*; Accepted *February 19, 1980*

Authors' Address: J. STEVENSON-HINDE, R. STILLWELL-BARNES, and M. ZUNZ: *MRC Unit on the Development and Integration of Behaviour, University Sub-Department of Animal Behaviour, Madingley, Cambridge CB3 8AA, England.*