

STUDIES IN INDIVIDUAL DIFFERENCES IN MAZE ABILITY

IX. RATINGS OF HIDING, AVOIDANCE, ESCAPE AND VOCALIZATION RESPONSES¹

ROBERT C. TRYON, CAROLINE M. TRYON, AND GEORGE KUZNETS
University of California

Received January 21, 1941

The complex emotional reactions of animals have been sorely neglected as the object of experimental study by comparative psychologists. Probably one reason for this neglect is that the comparative psychologist, being like everyone else a product of his culture, and living in an era of efficiency, sees as important the study only of the mechanics of skill in learning and problem-solving. Another reason is that he has been largely interested in the study of the behavior of the average animal and in the experimental variation of conditions which make the average change, whereas the study of the emotions and "personality" largely place emphasis on individual differences, a field in the main ignored by comparative psychologists. When, as in our series of experiments, the main objective is the study of individual differences, then emotional differences between animals immediately becomes one focus of interest. We cannot help but note that the emotional aspects of a rat's behavior often shape and color in marked degree the nature and quality of his learning, so much

¹ The three collaborators are responsible for planning and executing the ratings, the first two for the analysis of the results, and the first for the write-up. We are indebted to Mr. M. Cazier, who served as assistant and who offered many helpful suggestions.

The writer wishes to express his thanks for financial assistance from the Carnegie Corporation of New York, the Research Board of the University of California, and the John Simon Guggenheim Memorial Foundation for statistical aid from the Works Progress Administration under Project No. 465-03-3-631-A2 at the University of California.

so that, as was shown in several previous papers (3, 4), these non-cognitive determiners are believed to be important parameters in maze learning.

Still another and probably the most important reason for neglect is the great complexity of emotional responses, which by their nature do not lend themselves to measurement by the piecemeal counting statistics that seem more adequate as an objective description of differences in learning ability. What we need are techniques to describe objectively, validly and reliably the "normal" variation between animals in the numerous types of dynamic emotional responses characteristic of a species. Most of the work reported has either been in the nature of biographical sketches varying widely in quality and of little force as a spur for further experimental work, or of descriptions of specially induced abnormal states, such as "experimental neuroses." The most notable exception is the work of Hall (1), who has discovered that fecal count appears to be a good measure of the normal range of differences in the excitability of rats. Without minimizing the value of Hall's fruitful researches, we nevertheless believe a real need exists for diagnostic methods which provide a more direct *psychological* description of numerous other types of emotional responses, and one which is at the same time an *objective* description of the nuances of emotional reactions of animals in a given situation.

We believe that the most promising universal technique of describing emotional differences is a judicious use of the rating scale. In adapting this technique to the study of animals, we find that the following general steps are necessary: (1) The observers, or raters, must devote considerable time in an initial exploratory period during which the animals are observed in more or less uncontrolled "natural" stimulating situations, the object being to discover what types of responses appear to be genuinely emotional, and what kinds of situations elicit these responses. (2) The situations must next be objectified in some describable experimental way so that one has assurance that the animals are to be compared in the same situation. (3) The observers must next decide specifically what general types of reactions are to be

rated, and for each type lay out a scale on which the animals are to be compared. (4) As a consequence of extensive practice in rating a preliminary sample of animals for a given type of reaction, the raters must develop identical concepts of the quality of behavior to be matched to each point on the rating scale.

If these steps are properly followed, it would appear that the rating technique opens up a rich field for systematic scientific study of complex "personality" reactions. The advantages of this type of study in animals over that in human beings is the obvious one, namely, that with animals we are in a better position to pursue more fundamental studies of the basic determination, conditions, and organization of personality. Using animals, we have greater control over factors which ordinarily confuse similar studies in human beings; we have greater assurance that the animals being observed are confronted with the same stimulating situation; we have greater control over the history of the subject before the time of observation; we have available subjects of different known hereditary stocks; and we can study more cases.

The object of this paper is to show in some detail the methods by which objective ratings of complex emotional responses in rats may be constructed. We shall show that, in a large sample of rats, the animals vary widely in certain emotional reactions, and that after training, raters can judge such differences with accuracy and nearly perfect agreement *inter se*. We shall show also the consistency of the observed differences over time and the interrelation of different types of emotional responses. Our major objective was to study the emotional differences between our maze-bright and maze-dull stocks, but though these animals were, in part, the subjects of our ratings, detailed comparisons of them will be reserved for a later paper.

EVOLUTION OF THE RATING SCALES

With a feeling of certainty that, if any future workers plan to develop rating scales similar to our own, they will trace our own somewhat fumbling steps, and to give a necessary background to an understanding of our final scales, we will present here the various stages of evolution through which our scales went.

The first consultation of the collaborators was characterized only by a unanimous conviction that obvious wide differences in the emotional behavior of rats existed, and that as the differences were patterned actions not to be described in piecemeal fashion some form of rating scale was indicated. In our selection of the types of "natural" emotional responses we planned to study (though we felt a real need for an exploratory observation of the rats in a great variety of situations) we decided to rest on the experience of seasoned rat handlers of the laboratory, who pointed out two general laboratory situations in which rats show genuine and variable "fear" responses. These two situations are first, being manually manipulated by a human being, and second, being confronted in an experimental situation with novel objects such as, in a maze, the doors, curtains, unfamiliar alleys, etc. Our conclusion was therefore to observe rats' responses in these two situations. The progressive stages and final results of our ratings in the first situation are given in this paper. The results of our ratings of the rats' responses to novel maze features will be presented in a later paper.

Having decided to rate emotional response to handling, we first proceeded to observe the reaction of a few animals to manipulation. These observations suggested two situations, in each of which we thought we could enumerate a check-list of response items. These we described as follows:

Situation 1: "Reaching for the rat, picking him up, then releasing him in the cage, waiting 15 sec., then reaching for him again." The rat's responses were listed as being of two sorts, *positive*, having the items, "comes to glove," "crawls in glove," and *negative*, with items, "runs from glove," "cringes in corner," "digs in sawdust," "squeals," "bites."

Situation 2 (following Situation 1): "Grasping and holding the rat for 15 sec." The positive responses were listed as "noticeable relaxation in hand," "can be turned on back," "tends to remain in hand on being released," and the negative, as "on grasping him, he squeals," "eludes grasp," "squirms in glove continuously," "fights to get out," "bites," "urinates," "defecates," "jumps on being released."

The next stage consisted of an effort to refine these situations and scale the responses.

In situation 1 we thought we could separate three types of emotional behavior:

- (1) A *hiding* response scaled in two degrees:
 0. "crawls under other animals, or food-can,"
 1. "vigorous burrowing in sawdust, or cringes in corner."
- (2) *Avoidance*
 0. "approach hand,"
 1. "neutral to hand,"
 2. "mildly negative to hand,"
 3. "extremely negative to hand."
- (3) *Avoidance vocalization* scaled in three degrees:
 0. "silent,"
 1. "squeaks only on being grasped,"
 2. "continuous squealing on approach of glove and on being touched."

In situation 2 we thought we could distinguish also three types of emotional behavior:

- (1) *Tension*, which later became *Escape*, scaled in four degrees:
 0. "lies motionless in glove,"
 1. "not relaxed, muscles tense,"
 2. "struggles occasionally, some attempts to escape,"
 3. "excessive, unorganized squirming; fighting to escape."
- (2) *Escape vocalization* scaled in three degrees:
 0. "silent,"
 1. "occasionally squeaks,"
 2. "fairly continuous squealing."
- (3) *Release* response from the hand was scaled:
 0. "tends to remain in hand,"
 1. "jumps on being released."

These crude rating scales went through six more critical revisions before the final eighth set of scales was decided upon. In trying out and modifying the seven preliminary sets we spent four mornings rating 50 rats. The changes came about largely from our increasing familiarity with the reactions of the rats.

After we rated each rat we compared our ratings. When we disagreed we discussed at length the concepts which led to our ascribing different scale values to the rat's performance. As the result of these discussions, we tried to come to an agreement on the quality of performance which should be identified by a given scale value. At the beginning we were able to distinguish only several degrees of response in a given situation. By the 50th rat, however, we came to the point where we thought we could discriminate and agree upon at least seven different scale points on each type of rating.

THE FINAL RATING SCALES

Table 1 presents the eighth set of five rating scales which we finally decided to use in our experiment proper. Note that each is a seven point scale having a short verbal description serving in a cursory fashion to describe the quality of performance at most of the points. We wish to emphasize that these verbal statements represent very inadequate descriptions of the quality of behavior identified by the scale values. Each scale point really refers to a *concept* of a certain behavior-pattern or quality as conceptualized by the rater, for which the scale number and corresponding phrase serve only as mere symbols. An objection that this is mentalistic is quite irrelevant; the important fact is that the raters were able, as we shall show later, to watch a given rat's complex behavior and agree almost exactly on the scale value to be attached to it. There is no doubt in our minds that any other experimenter who would go through our preliminary training would, in due course, make ratings which would match our own.

It is to be noted that some of the scale points are labelled "Intermediate." Such a point refers to a quality of behavior lying between those on each side of it, a quality for which, though we could identify it by number, we were unable to agree upon a description in words.

Intensive practice with the final scales

Though we had rated 50 rats on the seven scales preceding the final set, we decided that before we should employ the final scales

TABLE 1
The final rating scales

<i>H</i> Hiding	Remains in open, non-stereotropic 1	Stereotropic 2	Momentarily hides under, behind animals, objects 3	Intermediate 4	Continuously hides, etc., cowering 5	Intermediate 6	Frantic hiding or burrowing or crouching 7
<i>A</i> Avoidance of hand	Neutral to glove, stands still 1	Tendency to avoid 2	Avoids glove, definite attempts to pull out 3	Slight effort to pull out 4	Vigorous pulling out of glove, usually successful 5	Intermediate 6	Difficult to secure, runs, jumps 7
<i>E</i> Escape from holding	Motionless 1	Restless 2	Sporadic or weak attempts to pull out 3	Intermediate 4	Struggles fairly continuously or vigorously to pull out 5	Intermediate 6	Frantic struggles, fights 7
<i>R</i> Release from hand	Remains at hand 1	"Flows" out 2	Organized leaving of hand 3	Intermediate 4	Vigorous directed jumping from hand 5	Intermediate 6	Blind frantic jumping from hand 7
<i>V</i> Vocalization	Silent 1	Intermediate 2	Occasional squeaks 3	Intermediate 4	Pronounced fairly frequent squeaks 5	Intermediate 6	Continuous squealing 7

on our experimental population we should practice intensively with these final scales, for later we would have to have make our judgments and write them down instantaneously. For four successive mornings we rated random samples of rats until we had finally covered a new sample of 70 animals. During this practice period we continued to discuss our discrepancies, and at the end we were confident that we had reached a high point of skill in following the procedure and entering our ratings, and that further practice would not greatly improve our agreement. We found that we could rate two hiding, two escape, and the avoidance, release and vocalization responses in the total time of about 1 minute. Having now spent eight days on preliminary ratings of 120 rats, we were ready for the ratings of our experimental populations.

Procedure and controls of the final ratings

As the experimental procedure and controls of the stimulating situation are important in securing good ratings, a detailed description of our technique is presented below.

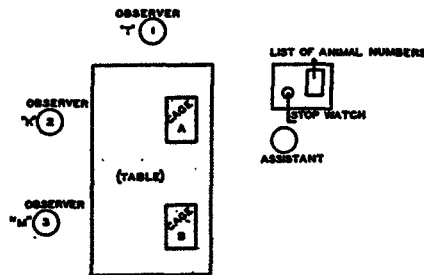
The lay-out. The disposition of the paraphernalia required for the ratings is shown in figure 1a. The three observers, raters *T*, *K*, and *M* are shown seated around the rating table, each having a clear view of cage *A* in which the rat being rated is located. Rater *T* is also the *handler* who manipulates the rat being observed. He manipulates the rat with his left hand, which is encased in a cotton glove that is thoroughly saturated with rat odors. In his right hand he holds his pencil ready to enter his own ratings on his sheet.

While the rats in cage *A* are being rated, cage *B* stands nearby on the table as shown. Thus the rats of cage *B* have about five minutes to become adjusted to the general situation while the rats of cage *A* are being put through. When cage *A* is finished, cage *B* is pushed into its place, and a new cage placed at *B*. Each cage holds four animals that have been living together for several weeks. Each cage is made of hardware cloth, having a width of 10", length of 14", and height of 9". The contents of the cage are a dry food-can so suspended inside as to leave about

an inch and a half clearance below it in which a rat may hide, a wet food-can for the daily ration, about an inch and a half of sawdust flooring, and four rats.

At the side is the assistant's small table, containing a stop-watch and a list of the rat's identification numbers per cage. The job of the assistant is to point out to the raters the rat to be rated, to call time, and to arrange the cages.

(A) POSITION OF CAGES, OBSERVERS AND ASSISTANT.



(B) DIAGRAM OF CAGE RATING PROCEDURE.

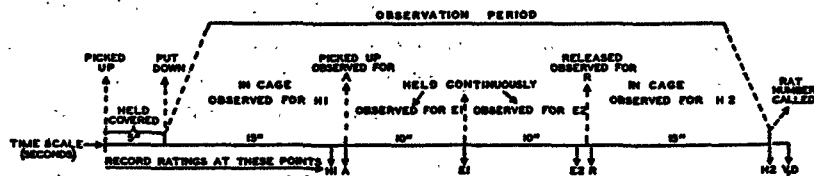


FIG. 1. DIAGRAMS OF THE RATING SITUATION (1a) AND OF THE RATING PROCEDURE (1b).

The procedure. Figure 1b is a schematic diagram of the procedure, indicating the passing of time as one reads from left to right. Above the line are the operations expressed from the point of view of the handler. On the base line is given the time intervals in which the observation for each type of rating is permitted. Below the base line is shown the point in time at which each rating is entered on the sheet. The successive steps are explained in detail below:

1. Looking on his list of animal numbers, the assistant finds the

rat in cage *A* to be rated, then looks in cage *A* until he locates the rat, and points him out to the handler, *T*.

2. *T* picks up the rat with the left gloved hand, holds him with his head in the palm, thumb and first finger over his back and under his forelegs. The rat is thus held with his head in darkness against *T*'s laboratory coat, *T*'s right hand over the rat's lower body. The rat is held firmly but not tightly, and usually lies motionless. The object of this holding is to provide a constant stimulating starting point of the ratings for all rats, and to present an identical handling situation to which they are to respond. As soon as the rat is correctly held, *T* says "Now" and the assistant starts his watch. From now on, the assistant keeps his eye on the watch and calls time routinely for each rating as indicated.

3. After 5 sec., the assistant says "Down," restarts his watch for the next 15 sec. of observation for the H_1 response.

4. *T* puts the rat down on his feet on the sawdust, preventing his having to jump, and removes the hand from the cage at once. The raters observe the hiding response. After 15 sec., the assistant says "Up," at which the raters enter their H_1 ratings on their sheets. During the hiding interval, a rat may crawl under his cage mates in such a fashion that one rater may lose track of him, in which case the rater immediately says "Which one?" and the other raters point out the rat with their pencils.

5. *T* reaches in the cage to pick up the rat, and the raters observe the avoidance response. *T* grasps the rat with the thumb and first finger under the forelegs, but with the head protruding out from the hand and away from *T*. If the rat is in a difficult position for this hold, *T* maneuvers him into position for the grasp. Every rat is touched first before being grasped. The *A* rating covers his response to being touched and grasped. In the case of an extraordinarily unruly rat, it may be necessary to grasp him by the tail, though this hold is to be avoided unless it is obvious that the tailhold is the only way to get him.

6. At the moment the rat in hand is lifted out of the cage, the raters enter their *A* ratings, and the assistant restarts his watch for the *E* ratings.

7. *T* holds the rat in a vertical position. The thumb is under the right foreleg, the arch between the thumb and first finger is over the rat's back, the first finger is under the left foreleg. The rat's head is thus pointed up, and his rear legs hang free. He is held sufficiently loosely to permit all escape movements to be expressed, but at the same time sufficiently firmly to prevent a fear of falling, or an actual escape. He is held about a foot and half above the table, and away from the cage or any other object that he can see near by. The objective is to rate his specific attempts to escape *from T* rather than his specific attempts to struggle *toward* some object. The raters observe his escape movements.

8. At the end of 10 sec., the assistant says "Ten seconds," and the raters enter their E_1 ratings.

9. *T* continues to hold the rat. After 10 more seconds, the assistant says "Down," at which moment the raters enter their E_2 ratings covering their observations of the rat's escape movements during the second 10 seconds.

10. *T* releases the rat into the cage as follows: He tips the rat forward, head about two inches above the sawdust, hind-quarters higher than the head, and then further loosens his hand but does not drop the rat, who must go out of the hand himself. When the rat lands on the sawdust, the raters enter their R ratings at once, and the assistant restarts his watch.

11. The raters observe the hiding movements of the rat for 15 seconds, at the end of which the assistant says "That is all," and the raters enter their H_2 rating.

12. The raters now enter the ratings on vocalization, V , based on their observation of the rat over the whole 55 sec. period.

13. If at any time during the period the rat defecates, the raters note the fact on their sheet.

14. The ratings now being complete, the assistant calls off the identification number of the rat, and each rater enters the rat's number above his ratings for the animal.

15. After the last rat in the cage is rated, that is, after every fourth rat, the raters compare their ratings, discuss discrepancies, but *do not change any of their ratings*. The object of this check is

to bring to the attention of any rater any shift in concept that is attached to each scale value.

It is obvious that considerable practice is required before final ratings should be attempted. The raters must be so familiar with the rat's behavior and with the scales that they can write down their judgments instantaneously. The handler especially must be thoroughly practiced, not only in this procedure, which requires double duty from him, for he must both handle the rats and rate them, but also he must learn the general "art" of handling rats that comes only with practice. Especially he must avoid sudden, jerky movements, and must learn just the right pressures required in grasping and holding the animals.

With respect to certain general experimental controls, the following items should be noted. In our experiments, on the first

TABLE 2
Types and numbers of animals rated for responses to handling

	BRIGHT (B)	DULL (D)	DULL BACKCROSS (BK)	HYBRID (M)	STOCK (S)	TOTAL
Males.....	40	36	13	65	23	177
Females.....	44	38	13	59	25	179
Total.....	84	74	26	124	48	356

day of rating a given group, the rats' cages were cleaned at 7 a.m., their daily wet mash being given at this time. At 9:00 a.m. the ratings began. On the second day of rating, the same controls were observed, except that the cages were not cleaned. The room in which the ratings were made was adjacent to the animals' living quarters, from which the rats' cages were carried directly to the rating table.

THE EXPERIMENTAL SAMPLE OF RATS

As we wished to compare the emotional responses of rats from our various stocks bred for differences in maze ability, we rated a large sample of rats. The types and number of animals rated are given in table 2. The 84 Bright and 74 Dull rats came from the F_{13} generation of our selective breeding experiment (2). The

26 Dull Backcross rats were the progeny of the F_1 of a cross between Bright F_{11} x Dull F_{11} , these F_1 backcrossed with their Dull F_{11} parents. The 124 Hybrid rats were the F_2 progeny of the F_1 mentioned just above, and they theoretically constituted a genetic random sample. The 48 Stock rats were a sample of a theoretically heterogeneous stock of rats from the Household Science colony at the University of California. Our total of 356 rats was therefore a very heterogeneous group of rats. These animals ranged in age from about 60 to 160 days old. None had been used in any psychological experiment, or had been handled more than was required at the times their cages were cleaned (about twice a week), at which times they were transferred manually from dirty to clean cages.

In addition to these 356 rats, we also rated 79 rats that had run the 17-blind T-maze. This latter group is not included in the analysis of results reported in this paper.

RANDOMIZATION OF GROUPS

For our ratings to be objective, it was necessary to be sure that the raters knew nothing about the rat being rated. Now, our practice with our hereditary strains has always been to mix up the progeny of various types of crosses in the living cages after the animals have been weaned and numbered with an ear-punch. In any cage presented for rating, there would thus be any combination of rats from the various strains. In view of the fact that the raters did not know the identification number of the rat being rated, they therefore had no way of knowing the type of animal being observed. But males and females were kept in different cages, and so were rats of significantly different ages. Furthermore, the Stock rats were not mixed in with our hereditary strains. To minimize any differences in ratings attendant on the raters knowing about these cage differences, the cages were drawn from the shelves in a completely random order, the assistant attending to these details. At any given time the raters never knew whether the next cage to be rated would be of males or females, young or old, Stock or hereditary strains. In view of this unsystematic variation, and especially considering that each rater

had to make seven ratings of complex behavior in a minute, the observers claimed that at the time the ratings were entered they completely ignored sex or age differences between rats.

This randomization of cages was instituted in each batch of cages rated at a sitting. The total group was rated in six batches of about 20 cages each. The first four batches were rated on the last eight days of April, 1935, the last two on July 6-9. Complete randomization could not be effected for the Hybrid group. As they were too young in April, they were rated only in July.

ACCURACY WITH WHICH EMOTIONAL RESPONSES ARE RATED

How closely do one rater's judgments of the emotional responses of the rats match the independent judgments of another observer? The facts are given in table 3.

The correlations between the ratings of one judge and those of another for each of the single types of emotional responses are given in subtable (a) of table 3. These values we may call the *raters' reliability*. Read the table as follows: in the first row of entries for the 177 males, the correlation of .85 represents the agreement of rater *T* with rater *M* on the first hiding response, H_1 , the score for each rat per judge being the mean of the two H_1 ratings on day 1 and day 2. Stated another way, rater *T* made two judgments of H_1 , one on day 1, the other on day 2, so for each rat we worked the rat's mean H_1 rating by *T*; the correlation between this mean rating by *T* and the analogous rating by *M* is .85, as shown. Similar coefficients, between *T* and *K*, and *M* and *K* are .81 and .84, respectively. It is important to stress the fact, which will shortly become evident, that the degree to which one rater makes an "error" which causes his mean rating to disagree with that of another rater's depends to a great extent on the number of ratings which enter into his mean rating. Thus we show in column three of the table the number of ratings entered into each score, *e.g.*, the coefficient, .85, is between the mean of *two* ratings of H_1 by judge *T* and *two* by *M*, or, briefly, 2×2 , as shown.

Scanning the values for the various types of emotional response, one sees that the agreements of these 2×2 ratings are quite substantial. This finding is surprising, in view of the fact that the

total time of observation of two ratings of each type was very short, being, for example, only 30 seconds for each of the hiding responses, 20 seconds for each of the escape responses, and often only momentary for the avoidance responses. The high agree-

TABLE 3

Raters' reliability: agreement between raters T, M, K on the mean rating per rat on two-day measures of each type of emotional response

(a) Single types of response

GROUP	RATERS CORREL.	NUM. RATINGS	HIDING		AVOIDANCE		ESCAPE		VOCAL. V
			H ₁	H ₂	A	R	E ₁	E ₂	
Males N = 177	T M	2 x 2	.85	.86	.88	.90	.88	.89	.96
	T K	2 x 2	.81	.85	.82	.83	.81	.88	.96
	M K	2 x 2	.84	.85	.83	.81	.84	.87	.96
	S-B	6 x 6	.94	.95	.94	.94	.94	.96	.98
Females N = 179	T M	2 x 2	.94	.92	.94	.92	.89	.91	.98
	T K	2 x 2	.90	.91	.94	.92	.86	.87	.97
	M K	2 x 2	.90	.92	.93	.91	.91	.87	.98
	S-B	6 x 6	.97	.97	.98	.97	.96	.96	.99

(b) Combined general types of response

GROUP	RATERS CORREL.	NUM. RATINGS	H ₁ + H ₂	A + R	E ₁ + E ₂	A + R + E ₁ + E ₂	H ₁ + H ₂ + A + R + E ₁ + E ₂
Males N = 177	T M	4 x 4	.90	.94	.92	.96 (8 x 8)	.95 (12 x 12)
	T K	4 x 4	.88	.89	.90	.93 (8 x 8)	.93 (12 x 12)
	M K	4 x 4	.89	.89	.89	.92 (8 x 8)	.91 (12 x 12)
	S-B	12 x 12	.96	.97	.97	.98 (24 x 24)	.97 (36 x 36)
Females N = 179	T M	4 x 4	.95	.96	.94	.97 (8 x 8)	.97 (12 x 12)
	T K	4 x 4	.94	.96	.91	.96 (8 x 8)	.97 (12 x 12)
	M K	4 x 4	.95	.95	.92	.96 (8 x 8)	.96 (12 x 12)
	S-B	12 x 12	.98	.99	.97	.99 (24 x 24)	.99 (36 x 36)

ments on vocalization are probably in part due to the fact that the judgments covered the two observation periods of 110 seconds.

If we set .95 as a minimal criterion of agreement, then these ratings between two observations per response are not satisfactory,

except for vocalization. But we have material here for determining the reliability of ratings based upon more than two observations. Actually, we do have *six* ratings on each rat for each type of response, those of the three raters on each of the two days. We can find the mean rating per rat of these six ratings; and estimate the rater's reliability of these total measures by the Spearman-Brown formula, provided, of course, we satisfy the assumptions of that formula, namely, that the judgments of all raters are equally good and interchangeable, *i.e.*, constitute comparable forms of the same "test."

We have internal evidence in the table relative to the equality of the judges. Note that for each type of rating, each judge agrees with the other two by about the same magnitudes. For a summary statistic on the matter, we have worked for each judge the average of his 28 correlations with the other two judges. The values come out as follows: For judge *M*, her average with the other two judges is .90, for *T* it is .89, for *K*, .88. These values are so similar as to indicate almost complete identity.

This equality of the judges in their accuracy of observation has special significance in this methodology of animal ratings. For the three of us judges varied widely in our experience with ratings and rats. Judge *M* has had years of training in rating social and personality traits of children, judge *T* has had years of experience handling rats, whereas judge *K* had had little experience of either sort. These differences in prior *general* experience had apparently been eliminated by the *specific* and extensive training on these particular ratings. It might, furthermore, have been expected that judge *T* would have been better, for he was the handler of the rats in the rating situation and thus may have received more data on each rat from the feel of the animals with his hands, yet this supposed advantage had no effect.

Having assurance of the equality of the judges, we worked the rater's reliability of the mean of six observations on each type of response, shown in the S-B rows of subtable (a). The value, .94, for H_1 of the males is, for example, the reliability of H_1 estimated by the S-B formula, $3r/(1+2r)$, where r is the mean of the three individual raters' reliabilities, namely, of .85, .81, and .84.

This value is interpreted as follows: the correlation, or rater's reliability, of the mean H_1 rating based on the six actual ratings of H_1 (two ratings each from three judges) with another comparable but hypothetical mean H_1 rating also based on six similar ratings, is .94. It is to be noted now that raters' reliabilities of all the types of emotional response, each based on six ratings, are close to, or above, .95, our minimal criterion of accuracy. For the females, some of the values are close to unity.

We wished to combine similar types of emotional response. Both H_1 and H_2 are hiding responses, A and R are two kinds of avoidance responses, and E_1 and E_2 appear to be nearly identical escape reactions. Furthermore, from the point of view of the general situation, the avoidance and escape responses are both immediate withdrawal reactions from the hand, and thus could be combined into one general index of such a reaction. Indeed, all six types of reaction when combined could be considered objectively as an index of general withdrawal from the presence of an interfering experimenter. As we shall later use such general measures, we naturally wish to know the magnitude of the raters' error in each of them. The rater's reliability of these combined measures is given in subtable (b). Note that the raters agree on these combined measures much more closely than on the single types of ratings. This betterment of agreement is of course due to the fact that each rater's score is a mean of more ratings. For the general hiding response, $H_1 + H_2$, for example, each rater's score is based on four ratings, i.e., both H 's on two days. For general avoidance of the hand, $A + R + E_1 + E_2$, each rater's score is based on eight ratings, and for the combination of all responses, each is based on 12. Many of the combined ratings show very high agreement between raters, meeting our minimal criterion of accuracy. By pooling the ratings of all three judges on each of the more general measures, we find raters' reliabilities that approach unity (the S-B values).

To sum up, our findings show unequivocally that under proper experimental conditions, the complex emotional responses of rats can be rated with complete objectivity and almost perfect agreement between trained judges. The amount of agreement is de-

terminated, in part, by the number of ratings entering into the mean score for each judge. We find that, for our single types of emotional responses about six ratings are necessary for minimal accuracy. We find that, given specific training our judges were equally good irrespective of prior experience with rats or ratings.

One outstanding finding, not to be smothered in the details given above, is *the proof of the existence of wide differences between rats in their emotional reactions*, even in very circumscribed situations. The raters' agreement in objectively describing or rating such differences depends to a large extent on such differences being very marked and obvious. Thus, the reason why the judges showed better agreement on females than males is that the former spread a great deal more in these emotional reactions than do males. For every one of these types of responses, the standard deviation of the females is greater than that of the males. Specific evidence on this matter will be presented in a later paper.

CONSTANCY OF INDIVIDUAL DIFFERENCES IN EMOTIONAL RESPONSES

In their responses to the same situation on successive days, how constant are the individual differences among the rats? Some evidence on this point is given in table 4.

For the single types of responses, the correlations between the rat's ratings on day 1 with those on day 2 are given in subtable (a). We will call these measures of constancy the *time reliability* of an emotional response. Read the table as follows: in the first row of entries, for the 177 males the mean rating per rat of the three judges' H_1 ratings on day 1 correlates .26 with the mean per rat of the same three judges' H_1 on day 2. Let us call these "raw" values the *time reliability for one observation*, meaning by this the correlation between a single measure on one day with that on another.

The one-observation time reliabilities are fairly low for the single types of responses. These facts signify that each rat tends to vary considerably in his responses from day to day. Part of the reason for this fluctuation is raters' error, for the raw values in subtable (a) are based only on three judges' ratings each day. But these raters' errors play only a minor role, as was shown in the

preceding section. A little of the unreliability is in the raters, but most of the inconstancy is in the rats.

What are the reasons for this fluctuation of a given rat's emotional response over two days? We can, of course, only speculate on the matter. One obvious reason is that our observations on a given day cover too short a time to sample adequately a rat's reaction. Recall that the $H_1 + H_2$ observation covers only a

TABLE 4

Time reliability: constancy of rats over two days of observation for each type of emotional response

(a) Single types of responses

	NUM. RATINGS	HIDING		AVOIDANCE		ESCAPE		VOCAL. V
		H_1	H_2	A	R	E_1	E_2	
Males								
Raw.....	3 x 3	.26	.40	.37	.69	.61	.60	.43
S-B.....	6 x 6	.41	.58	.54	.82	.76	.75	.60
Females								
Raw.....	3 x 3	.52	.61	.62	.72	.57	.56	.77
S-B.....	6 x 6	.68	.76	.77	.84	.72	.72	.87

(b) Combined types of responses

	NUM. RATINGS	$H_1 + H_2$	A + R	$E_1 + E_2$	A + R + $E_1 + E_2$	$H_1 + H_2 + A + R + E_1 + E_2$
Males						
Raw.....	6 x 6	.50	.61	.66	.69 (12 x 12)	.63 (18 x 18)
S-B.....	12 x 12	.66	.76	.79	.82 (24 x 24)	.77 (36 x 36)
Females						
Raw.....	6 x 6	.64	.72	.61	.72 (12 x 12)	.72 (18 x 18)
S-B.....	12 x 12	.78	.84	.76	.84 (24 x 24)	.84 (36 x 36)

total of 30 seconds, $A + R$ covers probably only a few seconds, and $E_1 + E_2$ covers 20 seconds. Were we to observe the rats more times in each of these situations on day 1 securing thus more generous samples of these reactions, we would probably have found much higher time reliabilities. A mean rating covering several such observations would tend to cancel out unsystematic "chance" emotional effects occasioned by adventitious variations in our technique of grasping and handling the animals.

Another unsystematic factor is the emotional effects on a given rat of the act of experimentally observing his cage mates. It was our feeling that the antics of one emotional rat tended to communicate the excitement to the others. If we happened to rate the most emotional rat in the cage first—one that showed aggressive avoidance responses—the other rats behaved more excitably when they were later rated than if we handled the aggressive rat last. In addition to these more observable factors, there are doubtless many other unknown components, including variable “internal” factors which cause the day to day unsystematic fluctuations.

The low value of some of these constancy coefficients may not, however, be due necessarily to “chance” variability of the rats on successive days. In their adjustment to a particular situation, the rats may *systematically* change in different ways. Some may become increasingly emotional, others decreasingly so, and still others may remain constant. In such a case, the correlation between ratings on different days might therefore be quite low, yet each rat show a systematic non-chance change in performance. A low time reliability, though truthfully recording an inconstancy of response, does not reveal the *cause* of the inconstancy. The best way to discover the degree to which the low value of a time reliability is due to “chance” variability of the rats and to varying types of systematic change among them is to make a detailed case study of each rat over many days of observing him in the same situation. One type of result would, however, show unequivocally that inconstancy was caused by unsystematic “chance” variations, and this would be the case in which the sigmas and the intercorrelations between successive daily observations were equal throughout. As we made but two daily observations we are unable either to make case studies or the statistical test, and hence we are unable to reveal the cause of the inconstancy in our rats, insofar as it exists.

For each type of response (for example, hiding) we will use later the rat's mean rating covering the two observations on each of two days, or a total of four observations. We need to know the correlation between this total measure and another comparable

measure of the same response also covering two days. The knowledge of the magnitude of this correlation, which is a reliability coefficient of our obtained two-day measure, is important, for it (or more strictly, its square-root) sets the limit to which our obtained measure could correlate with any of the other emotional response ratings. Though we lack a comparable measure of each response, we can, nevertheless, find the approximate value of the reliability coefficient from the Spearman-Brown formula, $2r/(1+r)$, where r is the obtained correlation between day 1 and day 2, i.e., the raw values in table 4. Thus, for hiding, the approximate reliability of the mean rating of four observations on two days is shown as .66 in table 4b in the S-B row. The value of .66 is the estimated correlation between our obtained two day four-observation rating and another theoretical comparable rating of hiding. We would thus not expect our obtained measure of hiding to correlate much higher than .66 with any other emotional rating, for it would correlate only to this degree with a theoretically comparable measure of hiding.

The reliabilities of the other types of combined emotional ratings covering four or more observations are shown in table 4b. The S-B values are quite respectable, being of order, .8. It appears, therefore, that when judges take at least four observations on each rat, the resultant mean rating provides a fairly stable emotional index of him. It would seem that the instability of the single observations, as revealed in table 4a is largely due to variable "chance" factors which tend to be annulled when more samples are taken.

ORGANIZATION OF THE EMOTIONAL RESPONSES

To the degree that each of these types of emotional responses is constant, are they all an expression of a common fundamental and general personality trait, say, of fearsomeness, or excitability? Expressing this question in quantitative fashion, if for each rat we secured a mean rating on each type of response based on enough observations to cancel out any raters' errors and especially to cancel the effects of "chance" variations of the rats, would the intercorrelations between such measures be unity? If such a

result did appear, the different types of response would be nothing more than different measures of exactly the same fundamental general emotional components. If, however, the intercorrelations were all zero, then each response would seem to be an expression of utterly unique and specific kinds of fundamental emotional components. If the intercorrelations were neither unity nor

TABLE 5

Interrelation between the emotional ratings of various types

The correlations for the males are above the diagonal, those for the females below.

(a) Single types of rating (6 ratings each over 2 days)

		H_1	H_2	A	R	E_1	E_2	V
	Rel.	.41	.58	.54	.82	.76	.75	.60
H_1	.68		.55	.16	-.03	-.01	-.06	.33
H_2	.76	.69		.15	.05	.17	.07	.27
A	.77	.51	.65		.65	.52	.40	.48
R	.84	.42	.65	.82		.71	.64	.41
E_1	.72	.34	.47	.59	.66		.87	.28
E_2	.72	.31	.42	.45	.60	.87		.20
V	.87	.53	.71	.82	.84	.54	.45	

(b) Combined types of ratings (12 ratings each over 2 days)

		$H_1 + H_2$	$A + R$	$E_1 + E_2$
	Rel.	.66	.76	.79
$H_1 + H_2$.78		.02	.02
$A + R$.84	.67		.71
$E_1 + E_2$.76	.45	.60	

zero, then it would look as if some common and some unique components were determining each response.

Our evidence on this question is presented in table 5, which gives the intercorrelations between the various kinds of emotional response. Read the table as follows: in subtable (a), the r of .55 above the diagonal at the intersection of H_1 and H_2 , is the correlation, for the 177 males, between the mean H_1 rating based on the two daily observations of H_1 by the three raters, and the

similarly composed H_2 rating. The analogous intercorrelation for the 179 females is below the diagonal. In the border row and column labelled "Rel." are shown the reliabilities of these same ratings taken from table 4. Recall that the raters' reliabilities of these measures, shown in table 3, are nearly unity, especially of the more general combined measures in subtable (b). The failure of the intercorrelations to be unity is thus not due to raters' errors.

Consider, first the hiding responses of the males. Looking at the reliability of H_1 , we note that this two-day measure correlates .41 with another theoretically comparable two-day measure of the same type of H_1 response. Similarly, H_2 correlates .58 with another like measure. Yet, H_1 and H_2 correlate with each other by approximately the same magnitude, .55. It looks very much as if H_1 and H_2 are expressions of the same type of fundamental emotional response, for they correlate with each other to the same degree that they correlate with repeated measures of themselves. A criticism of this logic is that, since the r between H_1 and H_2 is between two ratings taken on the same two days, whereas the reliabilities are estimations of correlation between ratings taken on different but theoretically comparable sets of two days, the r of .55 may be due exclusively to components overlapping H_1 and H_2 on each day. While this interpretation may be correct for H_1 and H_2 , such alleged daily components linking the two hiding responses do not act like *general* daily factors affecting the other types of avoidance and escape responses. Look at the correlation between the hiding responses and A , R , E_1 and E_2 ; these values are close to zero, yet they represent the relation between different ratings taken on the same days. The absence of relation between the hiding and the other responses is more clearly evident in subtable (b) where the r 's are of the order, .02 for males. This uniqueness of the hiding reaction relative to the avoidance and escape responses is all the more surprising in view of the fact that in our rating procedure the two hiding responses are temporally closer to the A , R , E_1 and E_2 than to each other, for H_1 and H_2 are the first and last ratings on each day.

With the females, the situation is somewhat different. Though H_1 and H_2 appear to measure pretty much the same kind of emotional response, their intercorrelation being .69 and their reliabilities not being much higher, .68 and .76, these hiding responses do show some relation to the avoidance and escape responses. This fact is quite clear in subtable (b) where it is seen that the combined hiding reactions correlates .67 with the avoidance, .45 with the escape responses. With the females there are either some common general emotional components at work in all the different types of responses, or if each type is determined by its own unique radical components (3), these radicals are correlated.

Turning to the A and R avoidance responses of the males, their intercorrelation of .65 is just about the mean of their reliabilities of .54 and .82, suggesting that they measure the same type of fundamental emotional reaction. But this avoidance response has a great deal in common with the escape reaction. Indeed, looking in subtable (b), the intercorrelation between these two general types of response is nearly as high as their time reliabilities. It appears as if avoidance and escape represent virtually the same fundamental kind of emotionality in males.

With the females, as with the males, the A and R avoidance reactions appear to measure the same type of emotional components, and though there is considerable overlap with the escape response, it is not as great as with the males.

With respect to the correlation between E_1 and E_2 escape responses for both males and females the r is of the order, .87, significantly higher than their respective reliabilities, which are of the order, .75. This increment over the reliabilities is probably due to adventitious daily factors running through the two measures, for it is to be remembered that in our rating procedure, they were merely two ratings of the rat while he was continuously held. This special daily overlap, amounting to the difference between .75 and .87, is probably due to special factors affecting the rat on a given day, but not appearing in different days and to a rater's "halo" resulting in rater's judgment of E_1 affecting spuriously, and quite irrespective of the rat's actual conduct, the judgment of E_2 . We have already noted that the escape responses of the

males appear to be fundamentally the same as the avoidance reaction, but with the females, though some positive relation exists, it is considerably less than the reliabilities.

The relations of vocalization, *V*, to the other types of emotional response are of great interest. Note that *V* shows in general the highest correlations throughout the table. Part of this general relation is due to an overlap of the situation in which *V* was rated with the situation of each of the other measures, for we rated *V* wherever it occurred in the rating procedure, whether in the hiding, avoidance or escape situations. *V* correlates highest with the avoidance responses *A* and *R*, but in comparison with the other situations it was in the avoidance situations that the rats did the most squealing. Note that, for the females, *V* correlates as high with *A* and *R*, namely, .82 and .84, as these two avoidance responses do with each other, namely, .82. It appears that fundamental components determining vocalization appears to be the same as those determining avoidance.

To sum up, it appears that the two independent ratings, *H*₁ and *H*₂, of hiding, measure the same fundamental types of emotionality, and similarly with the two measures of avoidance, and of escape. But these three types of emotionality are not necessarily identical. Among the males hiding appears to be a unique type, whereas avoidance and escape appear to be virtually the same. Among the females, the picture is more complex. The various responses all intercorrelate positively, suggesting as one possibility a general common factor of emotionality at work. But each of these types is nevertheless somewhat unique, for they do not correlate as high with each other as they do with repeated measures of their own type. Vocalization appears to reflect the same type of emotionality as avoidance.

A special word should be added about rater's "halo," a *bête noire* as a source of correlation between ratings of human personality traits. Except for a possible mild effect on the two escape responses, we find no convincing evidence of halo operating under our conditions of training and procedure. In the first place, "halo" could not operate to account for our time reliabilities, for on the second day of rating, as we never knew the identification

number of a rat until after he was rated, we had no way of remembering his previous day's rating. As we rated about 80 rats at a sitting, presented in a different order from that of the first day, any possibility of recalling a given rat from his position in series, was remote. Had halo operated in different ratings on the same day, one would anticipate the intercorrelations *between* ratings being higher than the time reliabilities in which halo could not operate. But except for the case of the escape responses, which are liable to halo, the intercorrelations are not higher than the reliabilities. Still another bit of evidence is relevant on this score. If halo were a factor, we can think of two ways in which it might be generated. One possibility is that the raters would get their general impression of the rat from his first behavior in the situation, namely, H_1 , and this would carry over to the later ratings. Were this so, then H_1 should show the highest correlations with the other ratings. The other possibility is that the general impression constituting the halo gradually emerges and approaches stability as the ratings proceed, in which case, the last rating, namely, H_n , should be the best index of the halo and should correlate highest with the other ratings. But for the males, both H_1 and H_2 are just the measures which correlate lowest with the other ratings, in fact, almost zero.

Defecation. Stimulated by Hall's work on defecation as a measure of emotional excitability (1), we planned originally to keep a record of each rat's defecations during the observation period, hoping later to investigate the relation of this more indirect physiological measure to our ratings. Under our conditions we found it difficult to secure an exact observation of this response, for we could only score it with certainty when the animals were initially held and during the escape ratings. During the hiding and avoidance reactions the rat ran around in the sawdust and defecation was, as a consequence, difficult to observe. The only record we kept was presence or absence of defecation during the entire observation period, but this turned out actually to be only during the time the animals were held. By the time we came to rate the final two batches of rats in July, we decided that, in view of the dichotomous character of this response and of our inability

in getting a complete measure of it, we would not make an effort conscientiously to record this response. For the first four batches of rats, we have, however, records on the defecations of 112 males and 120 females.

Figure 2 shows the relation between our direct ratings of emotionality and the indirect defecation index. Read the chart as follows: the first bar represents eight male rats that earned a mean rating of 2, covering the 12 observations of emotional responses,

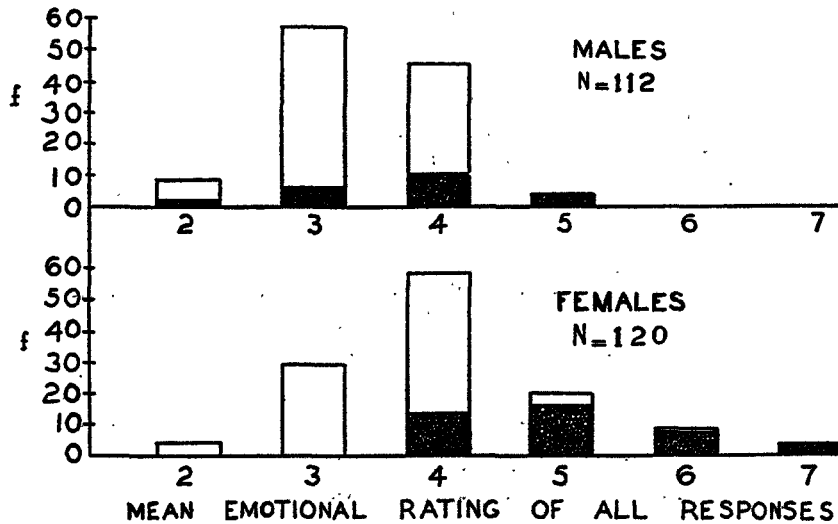


FIG. 2. RELATION BETWEEN MEAN EMOTIONAL RATING OF ALL MEASURES (H_1 , H_2 , A , R , E_1 , E_2) AND DEFECTION.

The ordinates are the number of rats, the abscissae are mean ratings, to the nearest integer. Black bars are number defecated.

namely, H_1 , H_2 , A , R , E_1 , E_2 , on two days. Of these eight rats, two defecated (black bar) on either day or both days during the observations. The next bar represents 57 male rats with mean emotion score of 3, and of these six defecated, etc.

The relation between rated intensity of emotion and defecation is very definite for the females. There are 31 females with the low emotion mean ratings of 2 and 3; none of these defecated. There are also 31 females with high emotion ratings of 5, 6, and 7; 29 of these defecated. The biserial correlation between defeca-

tion and emotion score is .83 for females. For the males the relation is obviously not as close, the biserial r being .34. The data do, in general, support Hall's contention that defecation is definitely related to emotional excitability.²

SUMMARY

Our aim in these experiments was to develop objective and reliable ratings of the complex emotional responses of rats. As the result of an extensive period of trial ratings on 120 rats we standardized a situation in which we later rated 177 males and 179 females for their hiding, avoidance, escape, and vocalization responses to handling. We found that, under our conditions, we could rate these responses with almost perfect agreement between the three different raters. These findings prove the existence of wide, observable differences between rats in these various types of emotional responses. The rats showed considerable day to day variability in a single observation of each type of response. Though such changes are probably in part due to genuine systematic changes of different sorts in different rats, we found, nevertheless, that the reliabilities of mean ratings based on four observations, of each type of response were of the order, .8. Our analysis of the intercorrelations between the various types of responses leads us to the conclusion that the emotional responses of rats are pretty complex affairs. The emotions of the male rats seem to be differently organized from those of the female. In males, hiding seems to be a unique response, avoidance and escape are closely similar, and vocalization and defecation seem to be mildly related to the other responses. In females, hiding, avoidance, and escape seem to be determined by some common general components, yet each has some uniqueness elicited by the particular situation in which it appears, and vocalization and

² Thus defaecation is a valid measure of emotionality, at least to the degree shown by the relation between it and our direct appraisal of emotional responses. L. I. O'Kelley (The validity of defecation as a measure of emotionality in the rat. *Journ. General Psychol.*, 1940, **23**, 75-87) has concluded that defecation is not valid. We would raise the question as to whether O'Kelley's indirect measures, with which defecation showed negligible relationship, constitute a criterion of emotionality.

defecation appear to be rather highly related to the other types of responses.

REFERENCES

- (1) HALL, CALVIN S.: Emotional behavior of the rat. *J. Comp. Psychol.*, 1934, 17, 89-108; 1934, 18, 385-404; 1936, 22, 61-68; 1936, 22, 345-352; 1937, 24, 369-375.
- (2) TRYON, R. C.: Genetic differences in maze-learning ability in rats. 1940, 39th Yrbk., Nat. Soc. Stud. Educ., Part I.
- (3) TRYON, R. C.: Studies in individual differences in maze ability. VII. The specific components of maze ability, and a general theory of psychological components. *J. Comp. Psychol.*, 1940, 30, 283-335.
- (4) TRYON, R. C.: Studies in individual differences in maze ability. VIII. Prediction validity of the psychological components of maze ability. *J. Comp. Psychol.*, 1940, 30, 535-582.