STUDIES IN INDIVIDUAL DIFFERENCES IN MAZE ABILITY

X. Ratings and Other Measures of Initial Emotional Responses of Rats to Novel Inanimate Objects[1]

ROBERT C. TRYON, CAROLINE M. TRYON AND GEORGE KUZNETS

University of California

Received February 13, 1941

That complex emotional responses of rats can be objectively and accurately rated by observers was proved in the preceding paper of this series (5). We presented there our findings on the ratings of responses of rats to being handled under standardized conditions, showing the requisite procedures for the securing of good judgments, and proving that rats do differ consistently in their emotional reactions of hiding, avoidance, escape and vocalization, and that these types of responses are complexly organized. Our general aim was to find relatively unrestricted, unartificial types of situations in which rats show a "normal" range of emotional responses. Extensive experience with rats had led us to choose the handling situation as being one in which emotional responses appeared.

But there is still another general situation in which we knew that significant emotional responses occur: this is the situation in which the rats for the first time confront the apparatus objects of an experimental situation and respond to them relatively freely.

[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, and for statistical aid from the Works Progress Administration under Project no. 465-03-3-631-A2 at the University of California.

The first time rats go into a maze, for example, they vary widely in their emotional reactions to new open alleys, to doors, curtains, corners, floors, etc. We decided, therefore, to develop a technique of rating their reactions to these novel maze features. It is the purpose of this paper to present our methods of judging these behaviors, to prove the existence of observable individual differences in them, and to show how these responses are related to each other and to the reactions to handling reported in the previous paper. The comparison of the emotional ratings of various strains of rats bred for different degrees of maze learning ability will be presented in a later paper.

As all the animals to be observed were to be run on our 17 unit *T*-maze, we had available a standard apparatus situation in which their emotional responses could be judged. We did not wish to judge their emotional reactions elicited during the actual learning of the complex design of this maze, for such responses would be complicated by emotional factors arising from differences in ability to cognize and learn the maze path, and thus would be a direct function of the very performance for which they had been selectively bred. We wished to rate their response purely to doors, walls, floors, curtains, etc., and we therefore looked for a situation in which the spacial path to food was so simple that all animals should be fairly equal in "knowledge" of the specific route to food through these maze features. In our standard maze procedure, we do, in fact, present the animals with just such a simple situation: before they enter the maze units proper, they are given eleven runs through a preliminary pathway, in which there are no choice-points and blind alleys, for the path leads directly to food. On the successive trials of this preliminary training, new doors, curtains, and sections of the pathway are gradually introduced to the animals. This preliminary training situation seemed to provide the type of situation we were looking for.

As the details of the general maze procedures have been presented elsewhere (1, 2), we need to consider here only the procedure of preliminary training. At the top of figure 1 is a diagram of the practice path. Each rat has an upper and a lower compart-

ment in the delivery "tub." When a given rat is made ready to run some of the later preliminary runs and even later, the maze

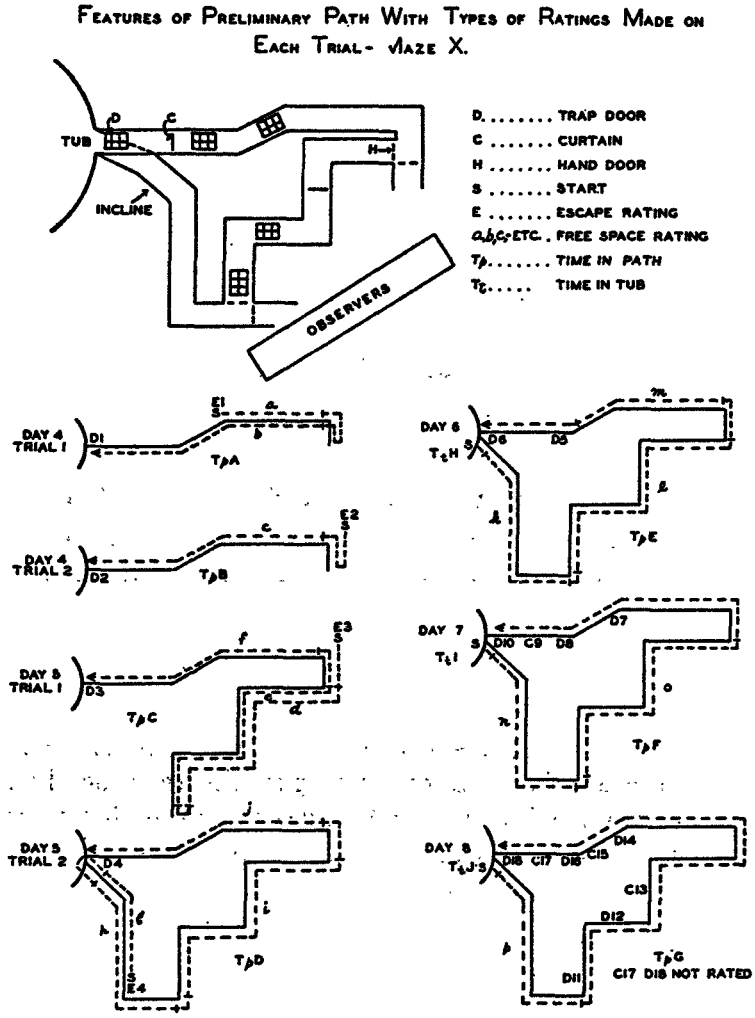

FIG. 1. DIAGRAMS OF THE PRELIMINARY PRACTICE PATH AND OF THE DAILY TRIAL RUNS, INCLUDING SYMBOLS REPRESENTING THE KINDS OF RATINGS AND OTHER MEASURES.

proper, he is placed in his lower compartment. When he leaves this, he climbs up the "incline" to the level of the maze floor, and

then proceeds around through the preliminary path to the tub again where he enters his upper compartment containing his daily ration.

During the eight days of preliminary training he does not, however, go through this complete circuit on all trials. On day 1, he is merely placed in his upper compartment, fed, and given 24 hours of adjustment to the tub. On day 2, he is taken out of his upper compartment twice, and given two runs of about 6 inches directly into his upper compartment from the very end of the practice path. The door, *D*, located at this point in the diagram lies flat on the floor and is not manipulated by the rat. When this door, and the others like it in the diagram, is not in use the elastic band, which normally holds it up at a 45° angle from its attachment on the floor, is released, causing the door to lie flat on the floor, where the rat can walk over it without interference. But when the door is up, the rat must walk upon it, causing it to sink to the floor. After he passes over it, it flops up behind him, thus acting as a trap door, preventing his retracing. On day 3, the door just before the entrance into the tub is drawn up, and the rat is placed twice before it, and each time he must go over it to his upper compartment and food. The remaining seven practice trials on days 4, 5, 6, 7 and 8 are diagrammed in figure 1. Each trip which the rat is required to make is shown by a dashed line, beginning with *S*, the start, and ending in the tub at the position of the arrow. All doors lie flat unless indicated by the symbol *D*. Note that there is a hand-operated door, symbolized by *H*, which permits us to shut off a section of the path. Black flannel curtains, which hang from the top and which completely block off the view into the path ahead, are introduced in later practice trials, where they are symbolized by *C*.

Scanning through these diagrams, one will notice that the doors and curtains, *D* and *C*, are followed by numbers. These numbers indicate the order in which the rat confronts these features. Thus from day 4 up to and including day 8, he is required to manipulate 18 doors or curtains. It should be mentioned that the rat is introduced into the practice path by hand on the four trials of days 4 and 5, but on days 6, 7, 8 he starts from his lower compartment by himself.

On looking over this practice procedure, the collaborators finally decided on certain situations within it in which we could rate or directly measure the emotional reactions of the animals, as follows:

(1) The rat's *escape* reaction on the four trials of days 4 and 5 when the assistant holds the rat in hand just before introducing him into the path. These are E_1, E_2, E_3 and E_4 just before S in the diagrams.

(2) The rat's *emotional reaction to doors and curtains* in the $D1$ to $D18$ situations shown in the diagrams. We decided, however, that these types of responses on days 2 and 3 are so brief, and so immediately tied up to response to the assistant's hand, and so difficult to see, that we could not properly observe the rat's behavior on these days.

(3) Emotional aspects of the rat's *movements through the pathway*. It happens that the practice path is made up of eleven portable segments like those out of which the T-units of the maze proper are constructed. The floor of each of these segments is balanced on a central fulcrum underneath in such a fashion that as the rat moves over the floor it dips under his weight, producing thus a slight rocking motion and clicking sound. In addition to the initial strangeness of alleys, these added stimulations would contribute, we thought, to variable emotional responses among the animals. There are certain sections of the rat's run that we decided to ignore. These are, in the main, those sections in which the rat has to turn around 180°. Note, for example, on day 4, trial 1, that the rat first proceeds to the east in the diagram, then turns around and finally comes back west to the tub. We decided that in the rat's turning movements, we could not separate those reactions having an emotional connotation from those gross movements required in turning, so we declared such sections to be "dead spaces" of his run in which we would ignore his behavior. In the diagrams these dead spaces are blocked off by vertical lines in the dashed line which symbolizes his whole run. Note that the sections of his run which are straight-forward are labelled $a, b, c, \cdots, p$. Such sections we called the *free spaces*, and it was only in these areas that ratings of emotional aspects of his actual running movements were to be made. Looking through the

diagrams, one will note that there are 16 such free spaces, labelled from *a* to *p* in the order of experience of the animal.

(4) The rat's *speed of response*, which we thought to be indicative of emotional reaction. The first of these is *tub time*, or starting speed, being the amount of time the rat takes on days 6, 7, and 8 to come out of his lower compartment into the incline leading up to the path. The hesitant and indecisive action of many animals at this point rather indicates that in many cases slowness in starting is indicative of fear of the situation. The three measures of tub time are symbolized by T_tH, T_tI, and T_tJ, in figure 1. The second speed measure is *total time* of running in the path from the start to finish on each trial. There are seven measures, T_pA to T_pG, for the seven days. Individual differences in time spent in the path, we thought, should certainly be in a large part a function of delays and hesitancies occasioned by fear of the general situation, though to a certain unknown degree sheer running ability unincumbered by fear would also be a contributing component.

EVOLUTION OF THE RATING SCALES

Having thus a promising standardized experimental situation available, we faced the problem of developing a series of rating scales which would objectively describe these "personality" reactions of the rats in it. We wanted ratings of the rat's emotional and motivational responses of fear, excitability, nervousness, exploration, vigor, etc. At the beginning, we were rather vague as to what we specifically were to observe for we did not know exactly and definitely what rats do in such a situation. From our experience with the ratings of responses to handling, we knew that the first step should be actually to observe in an exploratory way a large preliminary sample of animals in the practice path and to discover, if possible, what are the animal's significant reactions in this general situation. We knew that, as before, during this exploratory period our rating scales would undergo numerous changes until we developed a final set of scales. Each scale would refer to a significant general type of response, and each point on the scale would symbolize a concept of

a pattern or quality of behavior belonging to the general type.

We therefore placed an observation platform on top of the maze proper, from which, with the aid of mirrors hanging above the preliminary path as an auxiliary aid, we observed the behavior of a sample of animals. For a number of days, the three judges, *T*, *M*, and *K*, studied and discussed the behavior of 60 rats who were introduced into the path by the assistant and who ran an assortment of the different practice trials as diagrammed in figure 1. A brief description of the evolution of our ratings from the first crude set up to the final set is given below.

We saw at once that we could rate, as earlier, the *escape*, *E*, *release*, *R*, and *vocalization*, *V*, responses to handling by the assistant just before the animal is introduced into the path on the four trials of days 4 and 5. But as response to handling was not to the point of our interests here, we ultimately gave up *R* and *V*, the latter especially because of its infrequency, but we kept *E*, using the same scale described in the previous paper.

We next considered a rating of *exploratory* reactions, characterized by the rat's rearing up on his hind legs and sniffing at walls, tops, doors, curtains, etc. We thought we could measure this by tallying the number of "off-floor sniffings." But then it became obvious that, as between rats, we would have to correct for total time spent in the path, using, say, such a measure as "number of sniffings per 10 seconds of run." As such a measure appeared too difficult to time, we finally gave it up.

Another significant type of response was the appearance of unadaptive reactions which we called *tics*. These were characterized by face-washing, scratching, head-weaving, body-shaking, teeth-chattering, all of which seemed, we thought, to occur most frequently in rats who showed greater general unadjustment to the apparatus features. But here again, since we could not control the time of the rat spent in the path, a correction for this variable would be required. We therefore abandoned this measure, and with some regret, for we felt that tics were significant reactions of the animals, being definitely tied up with their adjustment to the situation.

Individual differences in *vigor* of movements appeared to be an important aspect of the rats' behavior. Our first scale of this type of response read as follows:

0. "continuously moves ahead."
1. "fitful stops, but goes ahead immediately."
2. "occasional stops of 10 sec. or more but goes ahead finally."
3. "prolonged sitting, has to be moved by experimenter."

As we went ahead with this scale, it became obvious that what we were primarily estimating was speed of the rat's movements through the path. We were not getting much more out of this scale than what was measured by time, T_t. What we wanted in our ratings were distinctions between the rats in their patterned emotional adjustments to the situation.

We now decided to set up two scales, one to rate emotional adjustment to the free space in the paths, the second, adjustment to the doors and curtains, endeavoring, in each case, to get distinctions between those patterned behaviors which were now becoming more apparent to us.

The first scale, *negativism in free space*, ran as follows:

0. "vigorous, undisturbed, positive, organized movement forward."
1. "some disorganization, confusion in movements."
2. "many tentative, confused, uncompleted movements."
3. "sits down, resists being pushed, cowers."

Negativism to doors or curtains had the following scale:

0. "no hesitation, goes through door or curtain."
1. "some hesitancy, may withdraw once."
2. "marked hesitancy, withdraws several times."
3. "put through by assistant."

THE FINAL SCALES

Finally, we came to the conclusion that we could best use one universal seven point scale to rate the emotional responses of the rats both to free space and to doors or curtains. This, the N, or *negativism*, scale is shown in table 1, which also contains the E scale used in rating of the escape response.

As with the rating scales reported in the previous paper, we must emphasize that the verbal descriptions at each scale point of the *N* scale do not do justice to concepts which these words attempt to portray. The phrases with their identifying numbers are mere symbols for the raters' concepts of qualities of the animal's performance. The observers required careful study of 60 rats before the concepts became crystallized, and the crystallization process came about by continuous discussion, rating, and cross-checking between the observers. We are confident, however, that any other observer would form closely similar concepts were he to go through the training period we underwent. The *N*

TABLE 1

Rating scales for the judgment of emotional reactions in the maze practice situation

Negativism, *N*, to free space, and to doors or curtains

No hesitation, goes through	Stops, investigates, no hesitancy	Positive response, stops, investigates, proceeds hesitantly	Intermediate	Negative response, withdraws or cowers, proceeds with extreme caution	Intermediate	Extreme negativism, put through bodily by assistant, continued resistance
1	2	3	4	5	6	7

Escape, *E*, from being held

Motionless	Restless	Sporadic or weak attempts to pull out	Intermediate	Struggles fairly continuously or vigorously to pull out	Intermediate	Frantic struggles, fights
1	2	3	4	5	6	7

scale was used to rate the responses of the animals to the 16 free spaces and 16 doors or curtains as represented in figure 1. Actually, the qualities of behavior represented by the seven points of the *N* scale are not the same in the reactions to free space as in response to doors or curtains. In free space, the rat is reacting to open alleyways; with doors or curtains, he is reacting to inanimate objects which he must manipulate. Though the words and numbers for comparable points on the scale as applied to these two types of situation are identical, the patterns of behavior which were given identical rating values in the two situations are rather different in their dynamic organization.

The scale points labelled "intermediate" represent, as described in the previous paper, patterns of behavior which the raters could distinguish from patterns which border them on the scale, and which the judges could nevertheless accurately evaluate, but points for which they could agree on no brief verbal description.

As with all ratings of complex behavior, the judge must come, through experience, to be able to "size up" the significance of manifold movements. A nice example of the necessity to take in the total significance of the rat's movement was the case of the rats who, on confronting for the first time a curtain hanging in an alley responded to it by withdrawing from it. Whether this reaction deserved a rating of 5 or higher depended on many other aspects of his behavior, such as cautious stretching, sniffing, and other evidences of fear. Some animals, however, on encountering the curtain responded as if it were an *end* of the path which had newly been introduced. These animals might withdraw from the curtain, come back and go through it as soon as they discovered by manipulation that it could be gone through. Mere withdrawal itself was thus not a certain indicator of a quality of negativism deserving a rating of 5 or higher; it all depended on the manner of the rats, and in the latter cases, especially on their reactions to the curtain when they ultimately went through it. Though some may criticize as subjective and possibly anthropomorphic a technique which requires the observer to use his judgment, the saving circumstance in our case is that we three judges were able to distinguish these different types of response and without collusion *to agree* in the distinctions.

Ratings thus have the very great virtue of enabling one to take in many aspects of the rat's performance not possible in physical measurements. But they also have the advantage of enabling one *to exclude* aspects of performance which vitiate the appraisal of certain responses. We wished, for example, to exclude from our ratings of emotional reaction to free space the effects of sheer speed of running of the rat, for which we had a time measurement. Two rats, for example, might spend equal time in the path, yet one show many signs of negativism, the other none. Our ratings,

we believe, accomplished such exclusion, for we cautioned each other against letting the factor of speed contaminate our judgments. We had a kind of control against such influence by use of the technique of "calling back" our ratings to each other after every fourth or fifth rat. Any rater who began to change his conceptualizations, either by relating them to factors other than those on which we had agreed, or by failure to size up the whole pattern of the rat's behavior, would thus shortly be appraised of his defections. The changing of any written down rating as a result of calling back was, of course, taboo.

Our technique, we believe, thus served to exclude the operation of a vitiating factor of "halo." As said above, we agreed to exclude as a halo-producing effect the factor of speed in our ratings. We also agreed to confine our rating of response to a given door, curtain, or free space to that feature alone, and not to let our rating of his responses to immediately preceding features influence our succeeding judgments. And we agreed to judge the whole response of the rat throughout a given free space, and not to rest our rating upon one dramatic instance in a particular part of it.

Specific procedure and controls of the final ratings

The general layout of the rating situation is shown in figure 1. The three judges sat on the observer's platform as shown in the figure. On the first four trials of days 4 and 5, the assistant, who introduced the rats into the path, called their identification numbers, and timed their run, stood near position *S* shown in the lower diagrams of figure 1. At the beginning of the rating period the rats were in their upper compartment of the tub, where they had been since their last practice run.

These compartments come in *sections* of seven compartments each, there being eight such sections, accommodating thus 56 rats in the tub. In addition to these 56 rats, there is an auxiliary set of shelves housing 28 more rats in four sections. As soon as the first 56 rats are run, 28 of them are removed from the tub, and the auxiliary set of 28 are placed in the tub for their run. Thus a total of 84 rats are available for each sitting. To avoid the males "tracking" the females, our maze procedure requires all the males to be run first, then the females.

Specifically, the procedure on the successive trials was as follows:

Days 4 and 5. At the beginning of the rating periods on these days, the animals were taken from their upper compartments of the tub, each group of seven rats in a section being placed in a single cage. The cages were placed in shelves next to the maze. Then the cages were brought as needed to position *S* in the diagrams. The assistant worked with four cages at a time. As soon as two cages of rats were finished, two new cages were added, thus giving the animals in these new cages about five minutes of adjustment before their turn came. Though on the seven trials the twelve cages were presented for rating in a fixed order corresponding to the sections in the tubs, the rats within a cage were presented in a random order from trial to trial, thus preventing the raters on later trials remembering each rat from his prior position in series.

Before introducing a rat into the maze, the assistant held him up in his left hand for the 10 seconds of his *E* observation by the judges, being careful to hold him away from the maze, and to preserve the grip described in the preceding paper for this type of situation. At the end of the 10 seconds the assistant said "Now," at which time the judges entered their *E* rating at once, and the assistant placed the rat on his feet in the path at *S* and started the stop-watch. At the end of the rat's run, the assistant stopped the watch and called the total time of running and the identification number of the rat. The raters then entered these data on their sheets. The judges entered their ratings of the rat's response in the path at the very moment the rat finished reacting to a given door, curtain or free space. The ratings made on days 4 and 5 are indicated by symbols in figure 1.

Days 6, 7, and 8. Before the runs on these three days, the rats are placed in their lower compartments. This operation is accomplished by transferring whole sections from upper to lower tiers of the tub without handling the animals. Only four sections are placed in the lower tier at a time, thus preventing a too-long wait of the animals that run later. On these three runs the rat has to come out of his lower compartment of his own accord.

When a given rat's compartment came opposite the opening thus permitting him to enter the practice path, the assistant started the stop-watch, and as soon as the rat reached the top of the incline, the watch was stopped and the tub time, T_t, was written down by the assistant. The watch was restarted now for the remainder of the run, providing thus the path time measure, T_p. When the rat reached his upper compartment, the assistant called to the judges the two time measures and the rat's identification number. The judges had in the mean time entered their path ratings. If a rat failed to come out of his lower compartment, he was removed, placed in a cage, and after the other rats had finished, he was given the run diagrammed in figure 1 for day 5, trial 2, the ratings on him being, however, the required one for the practice trial being observed. The tub time for this type of animal was recorded as the time between his entering his lower compartment and returning up to the top of the incline. All of the ratings made on days 6, 7, and 8 are symbolized in figure 1. On day 8, the raters reported that the rats showed no significant differences in their responses to C17 and D18, the curtain and door just before entrance into the upper compartment, these obstacles eliciting little negativism. Ratings of response to these two obstacles were therefore abandoned on this last day.

With respect to general controls, the times of day when ratings were made were approximately as follows: For days 4 and 5, the first trial was run between 10 and 12 in the morning, the second, between 1:30 and 3:30 in the afternoon. On days 6, 7, and 8, the ratings began at 10 in the morning and lasted about three hours. The food ration placed in the upper compartment was wet Steenbock mash (2), 1 level teaspoon for each run on days 4 and 5, but for the single trial of days 6, 7, and 8, the males received 1½ level tablespoons; the females, being smaller, received 1 level tablespoon.

THE EXPERIMENTAL SAMPLES OF RATS

After our preliminary ratings of 60 rats, we proceeded to the ratings of our experimental sample, consisting of 121 males and

120 females. The types and numbers of rats comprising the total sample of 241 are shown in table 2. All but one *B* male were the same rats who had previously been rated in the handling situation. The origin of these strains was described in the preceding paper (5).

In comparing the present sample with the total sample rated in the handling situation, one will note that out of the original group of 356 rats, only 241 were rated in the practice path. Our discarding of rats was entirely conditioned by the needs of the selective breeding experiment, in which the rats of table 2 were specially chosen groups to run the maze. We eliminated entirely the original 48 rats of the Stock group, for they were not subjects in the inheritance experiment. From the original Bright and Dull

TABLE 2

Types and numbers of rats rated in the practice path

	BRIGHT (B)	DULL (D)	DULL BACK-CROSS (BK)	HYBRID (M)	TOTAL
Males.....	26	26	9	60	121
Females.....	23	31	12	54	120
Total.....	49	57	21	114	241

groups, we furthermore discarded 62 rats. These rats were in the main those of the older litters, for it is our routine practice to select to run the maze the younger litters from the available litters, other things, of course, being equal. Some of the eliminated litters were, however, animals whose parents were not extremely maze bright or maze dull as the parents of the selected litters. Six rats were eliminated whose physical condition was not good.

The rated group was a heterogeneous sample of rats. With respect to maze ability it was doubtless somewhat more heterogeneous than a random sample, and probably to a less extent more variable than a random sample with respect to emotional responses in the practice path.

RANDOMIZATION OF GROUPS

As with the ratings of responses to handling, we wished to randomize the groups throughout the rating situation. The total sample was rated in the practice path in three batches of about 84 each, beginning on June 6, July 17, and August 15, 1935. The Bright and Dull groups were distributed throughout all three batches. But the Backcross appeared only in the second batch, and the Hybrid group was distributed throughout the second and third batch. In each batch of 84 the rats from the different groups were distributed at random with respect to the order of running and rating, so the raters never knew from the order of presentation in what group a given rat belonged.

Randomization with respect to sex was, unhappily, not possible. In a given batch, the group of females always followed the males. The raters knew this systematic organization, so that if there were any "sex halo" affecting the ratings it had a chance to operate. The judges did, however, claim that considerations of sex never affected their ratings, and that their judgments were determined only by the specific responses of the rats in the situation.

RATERS' ACCURACY OF JUDGMENT

The data which show the accuracy with which the judges agree with each other in rating the emotional responses are given in table 3. Read the table as follows: in the first row of entries for the 121 males, the correlation of .90 represents the agreement of rater T with rater M on the escape reaction, E, which for each rat consists of his mean rating in the four situations, E_1, E_2, E_3 and E_4. In other words, each judge rated each rat in the four situations, giving thus a mean E for each rat as determined by that judge. The r of .90 is the correlation between the 121 E scores as determined by judge T and the 121 E values of judge M. If we let n stand for the number of ratings entering into a given mean rating on each rat, then this r is a 4 x 4 correlation, as shown in the n x n row in the fourth row of entries in the table. Analogously, each judge rated each rat in the 16 free space situations,

$a, b, \cdots, p$, diagrammed in figure 1, thus each rat possessed a mean F value per rater. The correlation between raters in pairs is shown in the second column of numerical entries. Analogous data for D are in the third column. The last column gives the agreement between judges for the mean responses of rats in all 32 situations within the practice path on all trials. In looking over the values, one notes that, except for the escape

TABLE 3

Raters reliability: agreement between raters T, M, K of the mean rating per rat on n measures of response to each type of situation

GROUP	RATERS CORREL.	TYPES OF SITUATION			
		Escape $E(E_1 \cdots E_4)$	Free space $F(a \cdots p)$	Doors—curtains $D(1 \cdots 16)$	Free space, doors—curtains $F + D$
Males N = 121	T M	.90	.94	.94	.97
	T K	.90	.93	.93	.95
	M K	.88	.93	.92	.96
	(n x n)	(4 x 4)	(16 x 16)	(16 x 16)	(32 x 32)
	S-B	.96	.98	.98	.99
	(n x n)	(12 x 12)	(48 x 48)	(48 x 48)	(96 x 96)
Females N = 120	T M	.87	.95	.97	.97
	T K	.87	.93	.96	.97
	M K	.86	.94	.96	.97
	(n x n)	(4 x 4)	(16 x 16)	(16 x 16)	(32 x 32)
	S-B	.95	.98	.99	.99
	(n x n)	(12 x 12)	(48 x 48)	(48 x 48)	(96 x 96)

response, the coefficients of agreement between judges approach closely to .95.

The failure of these agreement coefficients to be unity is caused, of course, by rater's "errors," and such errors in any mean rating per rat diminish as the number of ratings upon which the mean is calculated is increased. We naturally therefore ask: If for each rat we find his mean rating for, say, E, based upon the 12 observations of all three raters, what is the raters' error in this new mean rating? We can answer this question definitely if we find that the rater's error is of about the same magnitude for each of the

three raters. The equality of the judges does seem evident in the table, for in each column type of situation, no one of the judges seems to be a better or worse judge than any of the others, for no one of them seems significantly to agree better or worse with the remaining judges than does any other. As the ratings of the various judges seem to be quite interchangeable and comparable, we may apply the Spearman-Brown formula and discover the raters' reliability of the mean ratings of the three judges combined. These values are given in the S-B rows of table 3. The value, .96, for *Escape* of the males is thus the raters' reliability of mean E based on the 12 observations. Note that for the other situations, the reliabilities are nearly unity, from which we conclude that our actual F, D, and $F + D$ ratings based on all judges are affected by virtually no raters' errors at all. In other words, *individual differences between rats in these ratings are entirely due to differences in behavior and are not in the nature of subjective errors of judgments of the observers.*

VARIABILITY OF THE RAT IN REPEATED RATINGS IN A GIVEN TYPE OF SITUATION

To discover how consistently the rat behaves over repeated observations of him in the same type of situation, we need two mean ratings of him covering all the observations of him in the same situation. With the escape ratings, for example, we may find his mean rating on $(E_1 + E_3)$ as one measure, and his mean of $(E_2 + E_4)$ as the other. Note that this odd-even "split-halving" of his four successive E_1, E_2, E_3, and E_4 ratings result roughly in equating in the two split measures any systematic adaptation effects that occur, for such changes will effect one split measure approximately the same as another. Having two such measures for all rats, we may then find the correlation between them, the value of which will signify the consistency of individual response in the particular situation.

The securing of two comparable measures of the responses of the rats in the free space situations presents some difficulties. We want the two split measures to include, or sample, the same array of stimulating conditions. We have 18 ratings, $a, b, \cdots, p$,

and they differ from each other in a number of ways. In the first place, they occur on different practice trials, and thus are affected differently by any progressive systematic change in emotionality in the maze situation. We finally decided on the following split measures:

F_1 = mean rating of *a, c, e, h, j, l, n, p*

F_2 = mean rating of *b, d, f, g, i, k, m, o*

How these splits are composed with respect to the effects of practice is shown in table 4, subtable (a). Read this table as follows: in the first F_1 row, rating *a* occurred on trial 1, *c* on trial 2, *etc.* It would seem that our arrangement of the eight ratings in each split spans well the seven trials.

A second circumstance to consider is the spacial situation in which the 16 ratings were made. Perusal of the diagrams of figure 1 will show that there are six free space areas which differ in spacial locale and in the direction which the rat faces. These situations are labelled I to VI in subtable (a) of table 4, and the disposition of the eight ratings of F_1 and F_2 with respect to these situations is shown there. It would appear that the two splits are about as similar a set of samplings of these six situations as one could make.

Finally, it would be important to include in the splits an equal number of *first responses* to these six spacial situations. Ratings that are of first confrontations with a given situation are shown in bold-faced type. It will be noted that there are three of these first reactions in each measure.

Statistical evidence of the comparability of F_1 and F_2 appears to the right in subtable (a), where it may be noted that the means and sigmas of these two split measures per sex group are very similar.

Similar complications occur in the task of securing two comparable split measures of the 16 responses to doors or curtains. We finally composed the split measures as follows:

D_1 = mean rating of D1, D3, D6, D8, C9, D11, C15, D16

D_2 = mean rating of D2, D4, D5, D7, D10, D12, C13, D14

How these split measures are composed with respect to the seven practice trials, with respect to the five doors, I to V, and three curtains, VI to VIII, and with respect to first confrontations

TABLE 4

Composition of the "split-half" ratings which make up the combined mean ratings, F and D, showing the comparability of half measures with respect to the trials and situations they sample

(a) Free space responses, *F*

SPLITS	PRACTICE TRIALS							MEAN		SIGMA	
	1	2	3	4	5	6	7	Male	Female	Male	Female
F_1	*a*	*c*	*e*	*h, j*	*l*	*n*	*p*	2.2	2.1	.43	.49
F_2	*b*		*d, f*	*g, i*	*k, m*	*o*		2.3	2.2	.48	.46

SPLITS	SITUATIONS					
	I	II	III	IV	V	VI
F_1	a	*c, j*		e, *l*		h, *n, p*
F_2		b, *f, m*	d	*i, o*	g	*k*

(b) Responses to doors or curtains, *D*

SPLITS	PRACTICE TRIALS							MEAN		SIGMA	
	1	2	3	4	5	6	7	Male	Female	Male	Female
D_1	*D*1		*D*3		*D*6	*D*8, *C*9	*D*11, *C*15, *D*16	1.9	1.9	.42	.48
D_2		*D*2		*D*4	*D*5	*D*7, *D*10	*D*12, *C*13, *D*14	1.8	1.9	.50	.56

SPLITS	SITUATIONS							
	Doors					Curtains		
	I	II	III	IV	V	VI	VII	VIII
D_1	D1, *D*3, *D*6	*D*8, *D*16			D11	C9	C15	
D_2	*D*2, *D*4, *D*10	D5	D7, *D*14	D12				C13

with these features (each measure having four) is shown in subtable (b).

The results on consistency of the rats in the two split measures of samplings of responses in each of the general types of situations are shown in table 5. Read the table as follows: In the first "Raw" row of values for the 121 males, the value of .64, for

Escape, E, is between the two split measures $(E_1 + E_3)$ and $(E_2 + E_4)$, as represented at the head of the column, and this value is a (6 x 6) rating, since each split is a mean rating based on 6 judgments, two from each judge. Analogous values for the other types of situations are shown in the row. We also wished a measure of consistency over all the 32 free space and doors or curtains situations. The value, .88, is such an index, being the correlation between the two split measures of the total 32, namely, between $(F_1 + D_1)$ and $(F_2 + D_2)$. For the case of time

TABLE 5

Consistency (reliability coefficients) of emotional measures in each type of situation

GROUP		TYPES OF SITUATION							
		Escape, E	Free space, F	Doors—curtains, D	Free space, $F+D$	Tub time, T_t			Path time, T_p
		(E_1, E_3) vs. (E_2, E_4)	F_1 (a, c, e, h, j, l, n, p) vs. F_2 (b, d, f, g, i, k, m, o)	D_1 (1, 3, 6, 8, 9, 11, 15, 16) vs. D_2 (2, 4, 5, 7, 10, 12, 13, 14)	(F_1, D_1) vs. (F_2, D_2)	*HI*	*HJ*	*IJ*	(*A*, *C*, *D*, *F*) vs. (*B*, *E*, *G*)
Males N = 121	Raw (n x n)	.64 (6 x 6)	.81 (24 x 24)	.75 (24 x 24)	.88 (48 x 48)	.53	.37	.55 (1 x 1)	.67 (4 x 3)
	S-B (n x n)	.78 (12 x 12)	.90 (48 x 48)	.86 (48 x 48)	.94 (96 x 96)		.74 (3 x 3)		.80 (7 x 7)
Females N = 120	Raw (n x n)	.63 (6 x 6)	.84 (24 x 24)	.85 (24 x 24)	.91 (48 x 48)	.56	.47	.48 (1 x 1)	.69 (4 x 3)
	S-B (n x n)	.78 (12 x 12)	.91 (48 x 48)	.92 (48 x 48)	.95 (96 x 96)		.75 (3 x 3)		.81 (7 x 7)

in the path, T_p, there was a total of seven measures which were split as shown at the head of the column. For time in the tub there were but three measures, H, I, J, a number which prevents a suitable split into two sets, so the interrcorrelations between these three are shown. In general the consistency of behavior, especially of the ratings in the path proper, is fair.

The raw measures of consistency represent, however, only *half* the observations on each rat in each situation. We naturally inquire as to the extent to which the variability of the rats in

each type of situation adulterates or produces unreliability in the rats' mean ratings covering *all* the observations in each situation. Now, as each of the raw values is an r between two fairly comparable measures, we may validly apply the Spearman-Brown formula to discover the reliability or consistency coefficient of each of the total measures. These values are given in the S-B rows of table 5, the S-B formula being $2r/(1 + r)$, where r is the raw value for all measures except tub time, T_t, in which case, the formula is $3r/(1 + 2r)$, where r is the mean of the three raw values.

For free space and for doors or curtains, the rats' unreliability is not great, the consistency values being of the order, .90. For the combined measure, $F + D$, it is, indeed, relatively negligible, the coefficient being of the order, .95. But for escape, and for the time measures inconsistency is still substantial, since the coefficients are a little less than the order, .80.

We would conclude from these data that the F and D ratings in the path covering seven days of observation measure accurately fundamental emotional differences between rats, being negligibly affected by rater's errors (the S-B values of table 5), and to a more sensible but still small degree by daily variability or systematic change in the rats. Comparing these findings with those on the emotional responses to handling, we find greater consistency here. The reason for our present better results, as presaged in the preceding paper, is probably that in our ratings in the practice path we more generously sampled the rat's responses over more days of observation. The only data we have that are strictly comparable to that of the preceding paper are our findings on the escape responses. In this study we find that the total reliability of four E responses is .78 for males and females; in the preceding study, the reliability of the four E responses was almost exactly the same, .79 for males and .76 for females.

ORGANIZATION OF EMOTIONAL RESPONSES IN THE PRACTICE PATH AND IN THE HANDLING SITUATIONS

What do these various indices of emotional responses in rats have in common? Do they reflect a common general fundamental set of emotional components, or are they complex functions of

different unique sorts? How are the responses to novel maze features related to the emotional responses in the handling situation? The relevant data are given in table 6. In subtable (a)

TABLE 6

Correlations of emotional measures in the maze situation with (a) each other, and (b) with ratings of response to handling

In 6a the value for males are above the diagonal, for females, below. The *r*'s of tub time and maze errors and time are based on a few less than the total *N*'s of 121 males and 120 females. All time measures are expressed as logs of recorded time.

(6a) Correlations with each other

PRACTICE PATH MEASURES		ESCAPE, *E*	FREE SPACE, *F*	DOORS-CURTAINS, *D*	TUB TIME, T_t	PATH TIME, T_p	MAZE ERRORS	MAZE TIME
	Rel.*	.78	.90	.86	.74	.80	.97	.98
Escape, *E*	.78		.15	.18	−.13	−.22	−.13	−.15
Free space, *F*	.91	−.08		.64	.59	.84	.49	.54
Doors-curtains, *D*	.92	.14	.70		.32	.81	.43	.47
Tub time, T_t	.75	−.07	.77	.58		.42	.29	.29
Path time, T_p	.81	−.06	.85	.79	.69		.51	.56
Maze errors	.97	−.16	.40	.44	.41	.50		.86
Maze time	.99	−.14	.51	.56	.46	.60	.89	

(6b) Correlations with ratings of responses to handling

CAGE RATINGS		Males						
Hiding, *H*	.66	.02	.13	.00	−.09	.16	.29	.28
Avoidance, *A*	.76	.22	−.14	−.24	−.17	−.18	−.36	−.39
Escape, *E*	.79	.50	−.10	−.14	−.09	−.11	−.26	−.25
		Females						
Hiding, *H*	.78	.06	.04	−.02	.04	−.02	−.03	−.01
Avoidance, *A*	.84	.10	−.17	−.18	−.09	−.23	−.31	−.37
Escape, *E*	.76	.42	−.12	−.22	−.18	−.12	−.32	−.33

* The reliabilities of ratings to handling are time reliabilities on the total samples rated.

the intercorrelations between the ratings and other measures in the maze situation are given, the values for males being above the diagonal, for females, below. In subtable (b) the correlations

(separated by sex) between these measures and the earlier ratings in handling situations are presented.

In studying these intercorrelations, one must keep an eye on the reliabilities of the various measures, shown in the border cells. These reliabilities serve as limiting values of the intercorrelations, for one would certainly not expect to find the correlation *between* two measures to be much higher than the correlation of each of them with another comparable measure of its own type of response. Indeed, if we find a given intercorrelation to approach the values of the reliabilities of the two measures correlated, this fact would at once imply that the two measures are comparable measures of the very same fundamental type of emotionality.

In scanning the values in this table, certain general facts at once appear:

(1) The values for the males and females appear to be closely similar throughout, suggesting that the emotional dispositions elicited in the maze situation have approximately the same organization. This result contrasts with our previous results on responses to handling, where we found the organization for females to be different from that of the males.

(2) We find that, except for the escape response, the emotional responses in the maze situation show virtually no relation to those in the handling situation. The values are close to zero and are preponderantly negative in value, suggesting a very mild tendency for those "unadjusted" in the handling situation to be "adjusted" in the maze situation.

(3) The escape responses, both in the maze and cage rating situations are virtually independent of the other ratings, both showing mostly low negative correlations with other maze ratings. They are themselves related by the order, .5, but not as high as their reliabilities would permit, suggesting that to some extent they are unique in the two general situations. Furthermore, the escape response in the maze situation seems independent of hiding and avoidance, whereas we found that escape in the cage rating situation correlated .67 and .45, respectively, with these reactions for the females, though zero for the males. The

fundamental components determining escape responses thus appear to be rather complex, and of two types: those cutting across all the cage ratings for females, but across only escape and avoidance, and not hiding of the males, and another different set cutting across the two escape responses in the cage and maze situations.

(4) Responses to free space, and to doors, or curtains, and path time seem to be highly interrelated both for males and females. This fact may not be surprising, since these particular three are measures of different aspects of the same general dynamic response in a given spacial path. There appears to be some uniqueness in both response to free space and response to doors or curtains, for their reliabilities are of the order, .90, though the intercorrelations are of order, .67. Paradoxically, these two ratings correlate as high with path time as the reliabilities permit. This complexity is difficult to explain, and about all we can safely say is that these three measures reflect pretty much the same emotional dispositions. One may raise the question: Since the physical index, path time, measures approximately what the "subjective" ratings do, what is the use of employing the F and D ratings? Our answer to this question would be that the ratings, which cover the same performance as does the time measure, are direct psychological descriptions of the performance, and they are, strangely enough, *more reliable* measures, and furthermore each seems to reveal to a certain degree unique aspects of emotional expression which the single crude measure of time masks.

Time spent in the tub, T_t, is closely related to the above measures for females, in view of its reliability of .75, though it obviously displays some uniqueness. For the males it is considerably less related to the other ratings, though for both sexes tub time most closely resembles the responses in free space.

These four measures, F, D, T_t, T_p, appear to elicit a fairly general type of emotionality in the maze situation, of which time in the path, T_p, appears to be the best index. But we must emphasize that path time is a broad sampling of various types of emotional negativism which are separately measured by F, D, and T_t. These separate measures manifest complex overlapping of

emotional components, but each reveals some dispositions not shown in the others. Incidentally, our finding that path time correlates as high with our qualitatively judged estimations of emotionality as its unreliability permits fits in nicely with the claim made in an earlier paper (4) that speed of running in the true paths of the maze is primarily determined by noncognitive components.

(5) The relations between later performance in the maze proper (as measured by errors and time made on 18 trials) and the various types of emotional response are shown in the last two columns of table 6. One of the main purposes of showing these relations is to provide concrete evidence of the role played by emotional components in maze learning. In several previous papers (3, 4), the existence of these noncognitive components was inferred from the kinds of behavior the rats showed in the maze units proper. Note in table 6 that both maze errors and maze time correlate with the broad measure of emotional negativism to maze features, namely, time in the practice path, T_p, to the amount .56 for males, and .60 for females. In the previous papers, the noncognitive component, of which we have here direct measures, was called *adaptation*, the term referring to degree of adaptation to doors, walls, curtains, alleyways, etc. This particular component was inferred as importantly affecting learning directly only on the earlier trials of the maze. But indirectly its effect in producing individual differences in learning would persist throughout the whole period of learning, for the animals who are most fearful and unadapted to these apparatus features of the maze, called "fundaments" would be "emotionally blocked" and fail adequately to develop the spacial orientations, called "direction-sets" and "higher direction configurations," necessary to the learning the true pathway through the maze.

It is to be noted that maze errors and time are in general negatively related to emotional responses to handling, except for the hiding response of the males. Good maze performers tend to be negativistic in the handling situation, though they respond positively in their emotional expressions in the practice path; poor maze performers tend to be the reverse, tractable to handlings,

unadjusted in the path. These facts will be more evident in the comparison of the various maze bred stocks with respect to these emotional measures, to be presented in a later paper.

The organization of these emotional responses thus appears to be very complex. The rat's emotions, it seems, are closely tied up to specific situations. In the handling situation, the two direct responses to the hand, avoidance and escape, because they are direct responses to the hand, are related, but hiding, which is free of the hand, is unique in the males. The escape responses in the maze and cage situation are related because they are both direct responses to the hand, yet our correlation analysis shows that they are somewhat different, and this fact is probably due to a general cage vs. maze difference in the two situations, though the elapse of time may be a factor. The three path responses, F, D, and T_t, are highly related because they occur in the same path situation. Time in the tub, or starting time, shows a lower relation to the three path measures, but it is measured in a different situation, namely, in the tub, which is, however, dynamically related to behavior in the path. Maze errors and time are most closely related to the four emotional measures taken in the practice path, because the performance in the maze proper is measured in situational replicas of the practice path. The females seem, however, to show more generality than do the males, for both in the handling and in the practice path situations the intercorrelations between their emotional responses are higher than those of the males. But when the situations are grossly different, as between the practice path and the cage rating situations, the females show as much specificity of response as do the males.

SUMMARY

As a sequel to our previous ratings on the emotional responses of rats to handling, we developed rating scales of the responses of a sample of the same group of rats to apparatus features in a preliminary practice path of a maze. We developed rating scales of the animals' reactions to free spaces or open alleys, to doors and curtains. In addition, we rated the escape response of

the rats at the time he was introduced into the practice path, and we secured a measure of starting or "tub time," and one of time spent in the practice paths. It was found that the raters' agreement on the mean rating per rat covering seven observation trials for each type of response was nearly perfect. The rat's consistency in each type of response, as determined from the correlation between split measures of each type of response ranged from .75 to .95, being higher for the ratings than for the time measures. The interrelations between these emotional measures, and between them and the previous measures of responses to handling, were studied. The organization of all of these emotional indices was found to be complex. In general, the responses elicited in a given situation tended to be highly related, but those in different situations were relatively independent.

REFERENCES

(1) TOLMAN, E. C., TRYON, R. C., AND JEFFRESS, L. A.: A self-recording maze with an automatic delivery table. Univ. Calif. Publ. Psychol., 1929, 4, no. 7, 99–112.

(2) TRYON, R. C.: Studies in individual differences in maze ability. I. The measure of the reliability of individual differences. J. Comp. Psychol., 1930, 11, 145–170.

(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.

(5) TRYON, R. C., TRYON, C. M., AND KUZNETS, G.: Studies in individual differences in maze ability. IX. Ratings of hiding, avoidance, escape and vocalization responses. J. Comp. Psychol., 32, 407–435.