THE INHERITANCE OF MAZE-LEARNING ABILITY IN RATS[1]

EDWARD CHACE TOLMAN

University of California

The problem of this investigation might appear to be a matter of concern primarily for the geneticist. None the less, it is also one of very great interest to the psychologist. For could we, as geneticists, discover the complete genetic mechanism of a character such as maze-learning ability—i.e., how many genes it involves, how these segregate, what their linkages are, etc.—we would necessarily, at the same time, be discovering what psychologically, or behavioristically, maze-learning ability may be said to be made up of, what component abilities it contains, whether these vary independently of one another, what their relations are to other measurable abilities, as, say, sensory discrimination, nervousness, etc. The answers to the genetic problem require the answers to the psychological, while at the same time the answers to the former point the way to those to the latter.

But as far as the present investigation is concerned it must be admitted that only the most elementary answers of either sort have as yet been obtained. The preliminary problems of technique and method of attack have proved all important.

1

The problem, as originally formulated, was merely the very simple one of trying by selective breeding to establish two independent strains of rats—one "bright"[2] and the other "dull."[2] Eighty-two white rats (41 males and 41 females) plus a certain number who for one reason or another had to be discarded, obtained from the stock of the anatomy department of this university, constituted the initial or P generation. This initial group had wild blood in them due to one wild grey male that had been mated into the anatomy stock some six years previous to the beginning of this experiment. And among the near relatives of the animals chosen for this experiment were many grey, black, black-hooded and grey-hooded individuals. The animals chosen for this experiment, however, were all of them white. The age of these initial 82 animals, although not known exactly, averaged about ninety days. The animals were received in small lots of about 20 each and were tested as received in a maze made of $7\frac{1}{2}$ inch unpainted redwood, covered with chicken wire, of the shape and dimensions shown in figure 1.

In all the tests, both of these initial 82 rats and of their descendants, the general method of procedure was the same and as follows:

Feeding. The animals were given at least five days of preliminary feeding in the food box. In some cases, however, this preliminary feeding lasted longer. The food used was bread and milk with sometimes sunflower seed mixed in. The duration of the feeding was approximately twelve minutes and took place in the food box.

The two mazes. When any given batch was being run, the males were run in one maze and the females in another; the two mazes were of the same shape and construction and were placed similarly as regards the light, which was artificial. In successive batches the males and females were interchanged as regards the two mazes.

Experimenter. In all the cases reported Mr. Frederick J. Adams acted as experimenter.

[2] In the sense of ability to learn a maze.

Scoring. The animals were in all cases run 1 trial a day for ten consecutive days and obtained on each day a score for number of errors and time in seconds. They also obtained a score for the number of perfect runs achieved in the total 10 trials. An error was counted as an entrance of a full body's length or more into a blind *while going forward*, or a retracing of more than one

Figure 1.
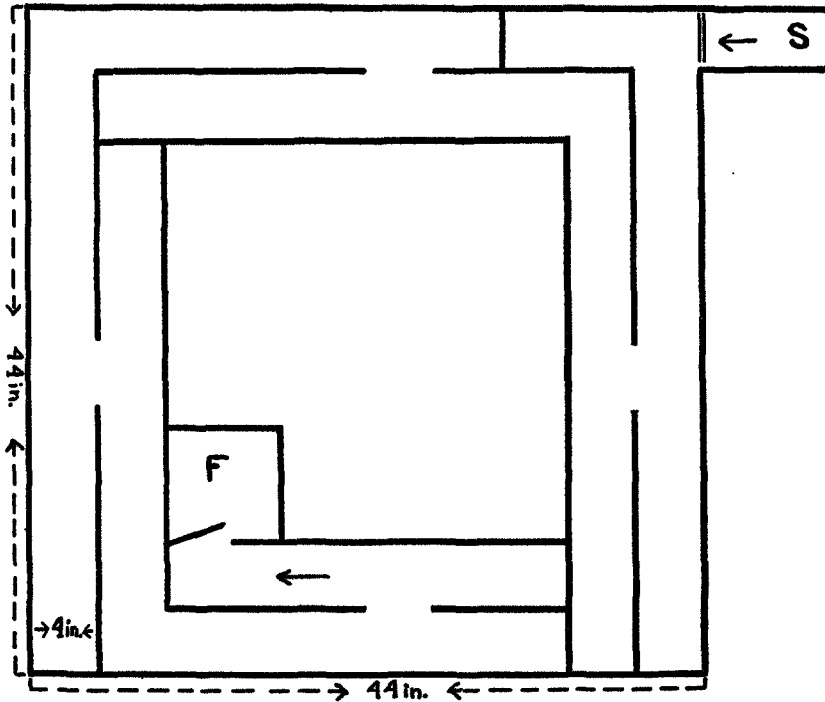
section of the true path —irrespective of how much more of the true path or how many blinds it might involve. The time was measured in seconds from the closing of the door behind the rat, after he had left the starting box, to the closing of the food box behind him, after he had entered the food box. These doors were manipulated by strings.

General plan of the investigation. Of the original (P) 82 rats, 9 supposedly bright males and 9 supposedly bright females were chosen to start the bright families; and 9 supposedly dull males and 9 supposedly dull females to start the dull families. The original 82 rats were run in successive batches of about 20 each and the two brightest males and the two brightest females and the two dullest males and the two dullest females were chosen in each batch. Some batches proved as a whole, however, brighter or duller than others, so that this method of having to select within a batch did not result in always picking the absolutely speaking brightest and dullest out of the total 82. The selection was based upon a rough pooling of the results as to errors, time, and number of perfect runs. In presenting our results, however, the data as to errors, time, and perfect runs will be given separately.

The family relationships among these first 82 rats were not known, so that in selecting bright males to be mated with bright females and dull males to be mated with dull females we may have been, in some cases, mating litter mates with each other. The F_1 generation was constituted by the litters from these original bright and dull matings. The F_2 generation was obtained by further selective breeding within the bright litters and dull litters of the F_1 generation, and in each case a brother was mated with a sister.

STATISTICAL TREATMENT OF RESULTS FROM ORIGINAL GROUP

In working up our data, we have considered our original 82 rats as an unselected group which could be used as a standard in determining the best statistical treatment for the remaining data. As Hunter and his students (4, 8, 9, 12) have shown, there is considerable question as to the possibility of measuring reliable individual differences in rats at all. And for that reason a discovery of the best statistical method of treating results was important.

Our first point of attack was to see what indications of reliability could be shown by correlating the error scores for different runs against one another. Table 1 shows the inter-

correlations (all the coefficients presented in this report are Pearson Product-Moments coefficients) for errors between successive trials 1 to 10 for all 82 rats.

The result is certainly discouraging. Apparently no two trials correlate with any degree of reliability. This is similar to the results reported by Hunter for time (9), only, if anything, even worse. In general, however, it may perhaps be said that the correlations near the diagonal, that is, between closely adjacent runs, tend to be higher than those at a distance from the diagonal, i.e., than those between far-distant runs. Hunter found a similar result, and his figures indicated it with even more certainty.

TABLE 1

Errors

n = 82

	2	3	4	5	6	7	8	9	10
1	0.145	−0.065	0.033	0.033	0.024	0.057	0.167	0.094	0.025
2		0.068	0.154	0.311	0.194	−0.002	0.019	−0.020	−0.122
3			0.116	0.109	−0.017	−0.129	−0.111	−0.099	−0.013
4				0.358	0.097	0.111	0.227	0.056	−0.068
5					0.123	0.209	0.172	−0.006	0.172
6						0.181	0.098	0.088	0.079
7							0.264	0.132	0.326
8								0.020	−0.020
9									0.305

Our next attempt was to correlate the *combined* error scores for a group of early runs against a group of late runs. Since the correlations between individual runs, and particularly those between early and late runs, was very low, we could hardly expect those between *combined* scores for the first part and the last part of the learning to be any higher. And such, indeed, proved to be the case (table 2).

Although no correlations (or even negative correlations) were thus obtained between the first and second halves of learning, it occurred to us that, if the results found by Hunter and somewhat suggested by our data in table 1 are correct—if, that is, the adjacent parts of the learning are nevertheless somewhat cor-

related—that better results might be obtained by combining alternate runs and then correlating odd vs. even. We therefore tried this out with the results shown in table 3. In certain cases, to save time, we obtained correlations for the 41 males, rather than for all 82, males plus females.

These correlations while unsatisfactorily low are at least better than those of the preceding tables. Furthermore, it will be noted that the highest correlation of all, 0.379, was obtained when runs 3 to 10, only, were included. In other words, it would appear that runs 1 and 2 were, as might be expected, very much

TABLE 2

Errors

n = 82

RUNS	COEFFICIENTS OF CORRELATION
2,3,4, vs. 5,6,7	−0.108
2,3,4,5, vs. 6,7,8,9	0.002
·2,3,4, vs. 5,6,7,8,9	−0.229

TABLE 3

RUNS (ODD VS. EVEN)	n = 41 (MALES) r	n = 82 r
1,3,5,7,9, vs. 2,4,6,8,10		0.276
2,4,6,8, vs. 3,5,7,9	0.344	0.296
3,5,7,9, vs. 4,6,8,10	0.179	0.379
3,5,7, vs. 4,6,8		0.324

subject to chance so that when included they tend to bring down the correlations. *We have, therefore, from this point on computed all our results from the data of runs 3 to 10 only.*

Further consideration of the whole situation suggested several possible treatments which might raise this correlation. Three different schemes were tried out. And, although we did not adopt any one of them in the actual working up of the further data of this report, we feel that they are of sufficient theoretical and methodological interest to be presented and to be kept in mind in the planning of future work.

The first one of these methods consisted in weighting the scores of the individual runs in terms of their respective standard deviations. Since the absolute numbers of errors are very much larger in the earlier runs than in the later, the adding of the crude error scores together means that the final result will depend much more upon the earlier runs than upon the later. Run 3, for example, will have a very undue weight in the total odd score, and so also will run 4 in the total even score. In order to get around this difficulty, therefore, we tried adding not the crude scores together, but the scores divided by the standard deviations of the respective runs. Thus the errors in trial 3 were divided by the standard deviation for errors in trial 3, and those in trial 4 by the standard deviation in trial 4 and so on. This caused the score in each run to have an equivalent effect upon the combined score.

The resultant correlation in the case of 41 males is shown in table 4.

TABLE 4

Weighted error score

41 males

RUNS	r
3,5,7,9 vs. 4,6,8,10	0.317

Comparing this with the correlation of only 0.179 obtained for the 41 males for the same runs (see table 3) when crude score was used, it appears that weighting does tend to raise the correlation, which means that a weighted error for all 8 runs would be a more reliable measure of the rats' error tendency than an unweighted score. The extra labor involved, however, is very great, and for the present we did not feel it to be warranted.

A second method of increasing reliability which suggested itself arose out of the observation that for the most part it seemed to be purely an occasional vagary in a few individual rats that brought the correlations down. The idea occurred to us, therefore, that in planning a future investigation it might prove desirable to discard all rats who in some one individual trial tended to vary excessively from the mean for that trial. A future

investigation might in short be limited to *consistent* rats only,
i.e., rats for whom the maze was really a reliable measure. In
order to see if this would work out, to see, that is, if we could
thus divide our rats into consistent and inconsistent, we tried
discarding from the original group all rats whose error score in
any single trial *varied two standard deviations* or more from the
mean for that trial. This resulted in our discarding 12 males
and 8 females. These discarded animals, it is to be noted, were
not ones whose *total* scores were necessarily widely variant from
the average; but merely rats who in some single trial or trials
(usually it was only one trial) tended to be variant from their own
average. The resultant correlations are shown below, not only
for odd vs. even, but also for 3, 4, 5, 6, vs. 7, 8, 9, 10 (see table 5).

TABLE 5

RUNS	29 MALES	62 MALES PLUS FEMALES
3,5,7,9 vs. 4,6,8,10	0.449	0.420
3,4,5,6 vs. 7,8,9,10	0.384	0.261

It is evident that the correlations go up somewhat for both odd
vs. even (compare with table 3) and also for first half vs. second
half (compare with table 2). The increase in correlation in the
case of the odd vs. even is not very great in the case of the total
(82). We conclude, therefore, that the method would have no
very great advantages to recommend it for future work.

Our third method of trying to get higher consistencies depended
not upon eliminating any rat's score, but rather upon trying to
iron out the meaningless excess variation in that score. When a
rat made an unusually great number of errors in any given trial
(due perhaps to excitement) or an unusually small number (due
perhaps to some lucky chance) such extremes did not seem to us
true measures of his real ability. Instead, however, of dis-
carding him altogether, we decided to try a method of scoring
which would *per se* discount those extremes—a method of scoring
which we could apply without fault of favor to all rats. This
method consisted in scoring simply in terms of greater than, less
than, or equal to the mean for each given trial. In other words,

his score for errors in any given trial was merely 1, 2, or 3 according as he made less than, equal to, or more than the average number of errors for that trial. The results for the 41 males and for all 82, trials 3, 5, 7, 9, vs. 4, 6, 8, 10, are shown in table 6.

These figures are the biggest of any we have yet obtained. Applying the formula[3] for determining the probable correlation on this basis of a score obtained from a whole maze learning (all 8 runs) against another whole maze learning, we would have as the probable correlations of the corrected scores for all 8 runs combined r = 0.699 for 41 males, and r = 0.675 for the 82 males plus females. These reliabilities are high enough to be almost satisfactory. One practical difficulty, however, in the application of this method to *our data as actually obtained* presented itself. The method would have been all right, if the data had been obtained so as to lend itself readily to working up into

TABLE 6

Corrected error score

RUNS	41 MALES	82 MALES PLUS FEMALES
3,5,7,9 vs. 4,6,8,10	0.537	0.509

terms of correlation coefficients. For, in such a case, the parents could have been scored in terms of their means and the children in terms of theirs. But in the present instance the experiment was so laid out that the chief results were to be found not in terms of correlations but in terms of comparisons of group averages (i.e., between "brights" on the one hand and "dulls" on the other). Such being the case, there was no common mean in terms of greater than, less than and equal to which it was reasonable to score these contrasted groups. As a matter of final procedure, therefore, we decided to reserve this corrected score as a possible method—for the future, and for the present to employ merely the crude unweighted scores.[4]

---

[3] Usually known as Brown's formula. $\dfrac{2r}{1+r}$. See Kelley, "Statistical Method" p. 206.

[4] In a minor investigation (3) which can be considered as more or less of an appendix to this report it was found as a matter of fact that although the corrected score method gave *apparently* much higher reliabilities within each individual maze, they did not raise the correlations between mazes.

We may turn now to a treatment of time scores. We did not attempt to try out all the different methods we used with errors— since the labor necessary seemed too great to be worth it. We tried merely the three methods of crude score, score weighted in terms of the standard deviations for the separate trials and score corrected in terms of above, below, or equal to the mean, for runs 3, 5, 7, 9, vs. 4, 6, 8, 10 and for the 41 males only. The results are shown in table 7.

It will be noted that here the crude score is just as good, if not better than, either of the others. Why this should be so we are not certain. A possible guess would be that the time scores are less given to extremes than errors because a rat tends to compensate. If, in other words, he is making an excessive number of errors, he tends to hurry up, and vice versa, if he is making few

TABLE 7

Time

41 males

Runs 3,5,7,9 vs. 4,6,8,10

Crude score	0.588
Weighted score	0.528
Corrected score	0.523

errors, he tends to take it more slowly. A careful study of time and error records side by side might throw light on this matter and must be undertaken in the future.

Before turning to our final genetic results, there is one more statistical treatment of the data of the original group to be presented. It consists in the intercorrelation between the 3 criteria, errors, time, and number of perfect runs.[5] The results are shown in table 8.

They come out as expected, save perhaps the zero correlation between time and perfect runs. To better understand this, however, we merely need to turn to the partial correlations, which are shown in table 9.

[5] No significant internal correlation for the number of perfect runs suggested itself.

The outstanding feature of these partials is that the partial correlation between time and perfect runs goes up to a respectable *positive* figure.   In other words, when the effect of errors on time has been rendered constant, it appears that the slower (more cautious?) rat makes the greater number of perfect runs, whereas the rapid (slap-dash?) rat continues to make a few final errors (due to too much haste).   This is a most interesting and important result and among other things it suggests a certain scepticism as regards any real significance in the time score.   For it would seem to mean that time is a measure of two quite independent factors.   On the one hand, it goes up in a purely mechanical

TABLE 8

n = 82

r_{ET}	0.515
r_{EP}	−0.624
r_{TP}	0.005

E = errors 3 to 10 combined (crude score).
T = time 3 to 10 combined (crude score).
P = number of perfect runs out of the total 10.

TABLE 9

$r_{ET.P}$	0.663
$r_{EP.T}$	−0.731
$r_{TP.E}$	0.487

fashion with errors—the more errors, the more time they take— but on the other hand, *per se*, it seems to be negatively correlated with errors, or at any rate with *final* errors.   It would seem, therefore, that the usual practice (5, 7) of favoring time as the one best single measurement is perhaps an unfortunate one.[6]

So much for the preliminary statistical treatment of the results for the P generation.   Let us turn now to the genetic problem, i.e., to the effects actually resulting from selective breeding.

[6] This suspicion of the time score was borne out in the results we obtained between the two mazes (3).   There, although the internal correlations within each maze by itself were high for time, the correlations between the two mazes were nil for time, although high for errors and perfect runs.

RESULTS OF SELECTIVE BREEDING

As before stated, we selected out of the original 82 rats of the P generation 9 bright males to be mated with 9 bright females, and 9 dull males to be mated with 9 dull females. The offspring of these matings constitute the F_1 generation, and fall into two distinct groups, "bright" and "dull," accordingly, as they are the offspring of the original bright matings or of the original dull matings. The F_2 generation was the product of brother and sister matings in the F_1 generation, a "bright" brother mated with a "bright" sister in the case of the "brights" and a "dull" brother with a "dull" sister in the case of the "dulls."

TABLE 10

Errors 3 to 10

	ORIGINAL UNSELECTED GROUP	"BRIGHTS"	"DULLS"	SELECTED TO BE BRIGHT PARENTS	SELECTED TO BE DULL PARENTS
	(n = 82)			(n = 18)	(n = 18)
P m	12.1			7.2	17.4
σ	15.6			5.1	8.6
		(n = 50)	(n = 59)	(n = 16)	(n = 13)
F_1 m		6.2	15.7	4.9	27.2
σ		3.5	13.9	3.0	22.7
		(n = 49)	(n = 39)		
F_2 m		10.5	16.4		
σ		4.9	7.7		

Not all the original 18 families (9 bright and 9 dull) continued till the F_2. This was due either to death or infertility.[7]

Two sorts of results have in general been computed and will be presented: one, group-means as between "brights" and "dulls," and two, correlations between parents and children in the cases of P_1 vs. F_1; and F_1 vs. F_2.

Tables 10, 11 and 12 present the former data, and table 14 the latter. The results for errors and time and perfect runs have been tabulated separately.

[7] Breeding has actually been carried on through 5 or 6 generations but with too small numbers to make the results altogether significant.

An examination of tables 10, 11 and 12 indicates that selection had the effect that the bright parents had in general very decidedly brighter children than had the dull parents. It

TABLE 11*

Time 3 to 10

	ORIGINAL UNSELECTED GROUP	"BRIGHTS"	"DULLS"	SELECTED TO BE BRIGHT PARENTS	SELECTED TO BE DULL PARENTS
	(n = 82)			(n = 18)	(n = 18)
P m	19.1			13.5	29.6
σ	11.9			5.4	25.2
		(n = 50)	(n = 59)	(n = 16)	(n = 13)
F₁ m		17.6	26.7	15.2	35.6
σ		13.9	24.1	6.5	36.2
		(n = 49)	(n = 39)		
F₂ m		17.5	19.5		
σ		6.1	9.9		

* The figures given in this table are in terms of 10-second units.

TABLE 12

Perfect runs—10 runs

	ORIGINAL UNSELECTED GROUP	"BRIGHTS"	"DULLS"	SELECTED TO BE BRIGHT PARENTS	SELECTED TO BE DULL PARENTS
	(n = 82)			(n = 18)	(n = 18)
P m	3.6			5.7	1.6
σ	1.5			1.1	1.1
		(n = 50)	(n = 59)	(n = 16)	(n = 13)
F₁ m		5.0	2.5	5.5	1.4
σ		1.5	1.7	1.3	1.4
		(n = 49)	(n = 39)		
F₂ m		3.9	2.2		
σ		1.7	1.8		

appears, however, that the further selection performed in the F_1 generation did not make the effect more pronounced in the F_2 generation. In fact, in all three tables the differences between the "brights" and the "dulls" is less in the F_2 generation than

in the F_1. And in each case it appears that this was due not so much to the coming up of the "dulls" as to the sinking of the "brights." How to explain this we are not at present sure. It seems possible that it may have been due to the strict inbreeding which was practiced in obtaining the F_2 generation. Our original stock was decidedly heterogeneous and inbreeding may well have uncovered some recessive taint as say, nervousness. This possibility seems to be borne out by such behavior tests as have been tried on Dr. King's inbred rats (2, 11, 13, 14). It is also possible that it was due to, or in some way affected by, the respective ages at which the animals were tested. The intention throughout was to test all animals at three months of age; but in the F_2 generation they tended to heap up on us with the result that some of them were tested late. This brought it about that

TABLE 13

Ages (days)

P (not known exactly but approximately ninety days)

	BRIGHT			DULL		
	n	m	σ	n	m	σ
F_1	52	96.6	14.3	55	88.9	13.4
F_2	49	108.6	16.0	39	124.5	17.5

they averaged somewhat older and this may have been part of the cause for the apparent sinking in ability of the brights (see table 13). To draw any certain conclusion upon this point, however, much further information would have to be obtained (e.g., the correlation and recession lines for age and maze-ability). Or, thirdly, it may have been due to something about our general living conditions which tended to cause our strain to degenerate, although at present we can think of no such cause. Our diet was practically the same as that of the original anatomy rats, i.e., garbage, save during the actual tests and for a couple of months during the summers. A future, better controlled investigation must settle between these possibilities.[8]

---

[8] The evidence obtained from our later generations, not here reported, from a group of brain weights which we measured, suggest that the matter was due to something more fundamental than mere age of testing, i.e., that it was either a matter of inbreeding or of nutrition.

We turn now to table 14 which presents the data for correlations of parents with children: P_1 vs. F_1 and F_1 vs. F_2. The two sets of variants in all these correlations consist of average figures for the parents on the one side, and average figures for their litters on the other.

In evaluating these figures it must be remembered that they result from arbitrarily selected groups and not from purely normal chance distributions. Of the 18 P vs. F_1 families, 9 were due to arbitrarily selecting the brightest individuals for mating and 9 from arbitrarily selecting the dullest individuals. Similarly of the 16 F_1 vs. F_2 families 8 were very bright and 8 very dull. In other words, we have only the two extremes of each distribution; and this would tend to raise the correlation from what it would be for the whole distributions. However that may be, the figures taken at their face value are of considerable interest in that they

TABLE 14

	ERRORS 3 TO 10	TIME 3 TO 10	PERFECT RUNS
P vs. F_1 (n = 18)	0.458	0.297	0.858
F_1 vs. F_2 (n = 16)	0.320	−0.267	0.779

pretty clearly substantiate our previous indications that time is an uncertain and ambiguous quantity. The time correlations are nil.[9]

And the second point of interest is that comparing the figures for errors and for perfect runs, the correlations for the latter are *much* the higher. One explanation that suggests itself is that since crude scores rather than corrected ones were used for computing the error figures, the error correlations were brought down beyond what might be conceived as their true or ideal values. The perfect run scores, on the other hand, were probably less affected by chance upsets, and hence more truly measured the real ability.

[9] Zero correlations have since been found for time, for one set of animals in two successive mazes (3).

A second possibility, however, is that perfect runs actually measure something which is really more truly and directly inherited than is that which is measured by errors. That is, our choice lies either between (a) assuming that the perfect run correlations between parents and children are higher than the error correlations, simply because the perfect run scores are in themselves a more reliable measure of one and the same thing that the error scores are a measure of, or (b) assuming that the perfect run scores really measure something different from the error scores and that this something different is more nearly inherited than is that which is measured by errors.

As between these two possibilities we shall not at present attempt to decide and it may well be that both factors play a part.

TABLE 15

	$n^P = 82$	$n^{F_1} = 109$	$n^{F_2} = 88$
r_{ET}	0.515	0.579	0.543
r_{EP}	−0.624	−0.671	−0.557
r_{TP}	0.005	−0.412	−0.179
$r_{ET.P}$	0.663	0.448	0.540
$r_{EP.T}$	−0.731	−0.582	−0.557
$r_{TP.E}$	0.487	0.038	0.179

Before concluding, we may give the intercorrelations between time, errors, and perfect runs for the F_1 and F_2 generations. We already have them for the P generation (tables 8 and 9). For purposes of comparison we will group them together for all three generations in table 15.

All the three generations present practically the same picture. The only exception is that in the F_1 generation the partial correlation between time and perfect runs when the effect of errors is eliminated is zero, not positive. With this particular group of rats, the effect of "cautiousness" seems not to have been very important.

Finally, in conclusion, we may point out that our results based upon more numbers substantiate those of Bagg (1). But

just how complicated a genetic trait maze-learning ability is, how many genes are involved, etc. —these are all problems for the future.[10]

Our main findings may be summarized as follows:

1. Practically zero correlations were obtained between numbers of errors made by the same rats in successive trials.

2. A correlation of 0.379 was, however, obtained between errors in trials 3, 5, 7, 9 vs. errors in trials 4, 6, 8, 10.

3. The highest internal correlation for errors was obtained from a "corrected score" in which the animal was scored in each trial simply in terms of whether he made more, less, or equal to the average number of errors for that trial. This correlation, trials 3, 5, 7, 9 vs. trials 4, 6, 8, 10 was 0.509 for 82 rats.

4. The intercorrelations between trials 3, 5, 7, 9 and 4, 6, 8, 10 for time was 0.588, and no higher for a "corrected time score."

5. From the intercorrelations of time, errors, and perfect runs with each other, it appears that time is an ambiguous measure that increases with errors, but at the same time increases with perfect runs.

6. Selective breeding in the P generation caused a very decided differentiation to appear in the F_1 generation between the descendants of "bright" parents and the descendants of "dull" parents. This appeared in all three scores —errors, time, and perfect runs. Further selective breeding caused no further differentiation in the F_2 generation. In fact, the differences tended to decrease. Whether this was due to inbreeding or to some extraneous factor we have not as yet determined.

7. The correlations between parents averaged and children averaged (P vs. F_1 and F_1 vs. F_2) were relatively high for perfect runs, low for errors, and very low, even negative, for time.

[10] At the present moment we are putting our energies upon the construction of some self-recording mazes five or six of which can be run simultaneously. This will very much lessen the labor involved and allow the carrying out of a much more extensive and better controlled investigation.

REFERENCES

(1) BAGG, HALSEY J.: Individual differences and family resemblances in animal behavior. A study of habit formation in various strains of mice. Archives of Psychology.

(2) BASSET, GARDNER C.: Habit formation in a strain of albino rats of less than normal weight. Behav. Mon., 1914, ii, no. 4.

(3) DAVIS, F. C., AND TOLMAN, E. C.: A note on the correlations between two mazes. Jour. Comp. Psychol. (Forthcoming.)

(4) HERON, WILLIAM T.: The reliability of the inclined plane problem box as a measure of learning ability in the white rat. Comp. Psychol. Mon., 1922, i, no. 1, pt. I.

(5) HIBBERT, HELEN B.: Time versus distance in learning. Jour. Animal Behav., 1914, iv, 60–69.

(6) HIBBERT, HELEN B.: The effect of age on habit formation in the albino rat. Behav. Mon., 1915, ii, no. 6.

(7) HICKS, VINNIE C.: The relative values of the different curves of learning. Jour. Animal Behav., 1911, i, 138–156.

(8) HUNTER, WALTER S.: Habit interference in the white rat and in human subjects. Jour. Comp. Psychol., 1922, ii, 29–59.

(9) HUNTER, WALTER S.: Correlation studies with the maze with rats and humans. Comp. Psychol. Mon., 1922, i, no. 1, pt. II.

(10) KELLY, TRUMAN L. Statistical Method. Macmillan, 1923.

(11) KING, HELEN D. Studies on inbreeding. II. The effects of inbreeding on the fertility and on the constitutional vigor of the albino rat. Jour. Exper. Zool., 1918, xxvi, 335–378.

(12) MAUPIN, OAKLAND: The reliability of the maze as a measure of learning ability in the white rat. (In preparation.)

(13) UTSURIKAWA, N. Temperamental differences between outbred and inbred strains of the albino rat. Jour. Animal Behav., 1917, vii.

(14) YERKES, ADA W.: Comparison of the behavior of stock and inbred albino rats. Jour. Animal Behav., 1916, vi, 267–296.