Applied Animal Ethology, 9 (1982/83) 299—315
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

GENETICS OF TRAITS WHICH DETERMINE THE SUITABILITY OF DOGS AS GUIDE-DOGS FOR THE BLIND

M.E. GODDARD[1] and R.G. BEILHARZ[2]

[1] *Department of Tropical Veterinary Science, James Cook University of North Queensland, Post Office, Queensland 4811 (Australia)*

[2] *School of Agriculture and Forestry, University of Melbourne, Parkville, Victoria 3052 (Australia)*

(Accepted for publication 9 February 1982)

ABSTRACT

Goddard, M.E. and Beilharz, R.G., 1983. Genetics of traits which determine the suitability of dogs as guide-dogs for the blind. *Appl. Anim. Ethol.*, 9: 299—315.

The most important traits causing dogs to be rejected as unsuitable for training as guide-dogs were found to be fearfulness, being too easily distracted, especially by other dogs, and aggressiveness. The guide-dog trainers evaluate these traits and several others using a series of 17 scores. The between-trainer repeatability of these scores varied from 0 to 0.7. Factor analysis of these 17 scores yielded 5 factors, which can be labelled distraction, general performance, sensitivity, fearfulness and fearfulness accompanied by high activity. There were no negative correlations between desirable traits, so it should be possible to obtain an overall improvement in the performance of the dogs. Comparison of dogs from the breeding programme of the Royal Guide Dogs for the Blind Association of Australia with dogs donated to the Association as puppies showed that the breeding programme had improved the dogs in the 3 important traits. Also, dogs reared under the supervision of the Association were superior to dogs donated as adults in these 3 traits. Females were more fearful and distracted by scents but less aggressive and distracted by dogs than males. There was significant genetic variation for fearfulness and possibly for dog distraction, suggesting that future selection on these criteria will further improve the standard of the dogs. These 17 scores, which are given early in a dog's training, have little ability to predict the dog's final performance on specific tasks but they do correlate with the overall reliability of fully trained dogs.

INTRODUCTION

The Royal Guide Dogs for the Blind Association of Australia provides and trains dogs as guides for the blind people. Initially, the Association relied on dogs donated as adults by the public but only a small proportion of these dogs proved suitable. In order to improve the way in which potential guide dogs are reared, the Association began a "puppy-walking scheme", in which pups are given to volunteers who raise them until they are old enough for training. Similar programmes exist in England and in the U.S.A. (Pfaffenberger, 1963). During this period, the puppy walker is instructed in raising the pup and

0304-3762/83/0000—0000/$03.00

visited regularly by a puppy-walking supervisor from the Association.

The Association also attempted genetically to improve its stock by selecting suitable dogs and bitches for breeding and breeding its own dogs. Thus both pups bred by the Association and those donated by the public are reared on the puppy-walking scheme. These dogs are returned to the Training Centre for training when 12—18 months old. The Association also accepts some dogs of the same age donated by the public.

Many guide-dog centres around the world have breeding programmes and some have claimed considerable success (Scott and Bielfelt, 1976). However, there is little published information on the success of these rearing and breeding programmes, or on the genetics of the traits which determine suitability as a guide-dog. This paper provides such information.

When the dogs are first brought into the training centre, the trainers spend about 3 weeks assessing them, after which they either accept them for training or reject them. Also at this time, they rate the dogs on 17 different scales which they believe measure traits which are important in guide-dogs. Since some dogs are rejected while on the puppy-walking scheme, the dogs which are scored by the trainers are a biased sample. Similarly, some adult dogs which are offered to the Association are rejected "at the front gate" and so do not appear in the records of the trainers' scores.

Because so many traits are important, the correlations between traits must be considered. If the 17 traits were uncorrelated it would be difficult to improve all of them simultaneously; or if 2 desirable traits were negatively correlated, improving one might make the other worse. This paper examines the correlations between these traits, and the genetic and environmental factors which influence them.

MATERIALS AND METHODS

The Association-bred dogs (A dogs) and dogs donated as puppies (P dogs) were placed on the puppy-walking scheme between 6 and 12 weeks of age. The males were castrated just prior to sexual maturity at 5—7 months of age. However, dogs donated as adults (D dogs) were still entire at the time the trainers scored them.

The 17 scores were each rated on a 0—5 scale with 0.5-point intervals. The scores were:

Nervousness (N) = fear, usually shown by withdrawal or inhibited movement, of people, traffic, strange places;

Suspicion (S) = fear mainly shown by approach — withdrawal conflict towards unusual objects;

Sound shy (SS) = fear of loud noises;

Anxiety (ANX) = fear shown in more subtle ways than nervousness, e.g. low tail position, unusual pulling on the lead;

Aggression (A) = threats or attacks directed at people or other dogs;

Nervous Aggression (NA) = aggression elicited by stimuli which also elicit fear;

Concentration (C) = attention to the stimuli to which the dog is being trained to respond;
Distraction (D) = attention to other irrelevant stimuli;
Dog Distraction (DD) = attention and attraction toward other dogs;
Cat Distraction (CD) = attention and attraction toward cats;
Nose Distraction (ND) = attention and attraction toward scents;
Willingness (W) = keenness to work and to carry out commands;
Hearing Sensitivity (HS) = strong response to sounds, including the trainer's voice;
Body Sensitivity (BS) = strong response to touch and leash corrections;
Temperamental Stability (T) = overall suitable temperament;
Initiative (I) = makes decisions, e.g. in traffic and in negotiating obstacles;
Excitability (EXC) = high activity.

The records of dogs scored between 1965 and 1976 were used. The total number of dogs for which records were available is 887 but some information was not recorded on all dogs, so the number in any particular analysis may be less than this. The scores EXC and ANX were added in 1972, so only 250 records are available for these scores. The number of dogs of each breed, source and sex is given in Table I. The distributions of several of the scores are skewed, with a tail on the undesirable end of the range. For analysis of variance, these scores were transformed to logs but the least squares means have been converted back to the original scale for presentation.

Beginning in 1975, 87 dogs were rated on the performance of their work when fully trained, in addition to their rating on the scores set out above. The trainer rated his or her dogs on 12 variables just before they were given to clients. The scores used were:

TABLE I

The distribution of dogs among breeds, sources and sexes

	No. of dogs		
	Total	Male	Female
Breed			
Labrador	731		
Golden Retriever	43		
Other breeds	37		
Unknown	76		
Source[1]			
A		189	146
P		38	139
D		209	166

[1] A = bred by the Association and raised on the puppy-walking scheme.
P = donated as a puppy and raised on the puppy-walking scheme.
D = donated as an adult and taken immediately into testing and training.

Traffic = ability safely to negotiate a path through road traffic;
Notices obstacles = reliability in guiding the handler around obstacles such as overhanging branches which the dog could pass through or under but which the handler could not;
Negotiates obstacles = ability to find a way around obstacles which block the direct path;
Right-shoulder work = reliability in leaving enough room on the right hand side for the handler to pass people, or to pass through doorways without bumping against them;
Places = ability to remember and find places which are visited regularly;
Ease of control = the ease with which unacceptable behaviour, e.g. dog distraction, can be prevented;
Stressful situations = ability to continue working well in stressful situations such as noisy dense crowds;
Strange places = ability to continue working well in unusual places which the dog had not encountered previously;
Speed = the natural speed at which the dog walks when working;
Learns quickly, retains well = speed of learning and reliability of retaining the tasks taught during guide-dog training.
Reliability = an overall assessment of the reliability of the dog's work.

These variables were scored on a 3-point scale: (1) satisfactory; (2) good; (3) very good (except speed which was scored as (1) slow; (2) medium; (3) fast).

RESULTS

The extent of agreement between trainers

One potential source of error in these scores is differences between trainers. Trainers could disagree in their scores in 2 ways. Firstly, 2 trainers might agree on the ranking of a group of dogs, but one trainer might give consistently higher scores than the other. Secondly, the 2 trainers might rank the dogs in a different order, i.e. the correlation between trainers might be low. The extent to which trainers rank dogs in the same order can only be assessed if 2 or more trainers score the same group of dogs. Normally, each dog is scored by only one trainer. Therefore, to provide information on this question, we arranged for several trainers to score the same group of dogs. This was done by 6 trainers watching 9 dogs given one walk each, all with the same handler. Consequently, the scores obtained are not completely comparable with normal scores which a trainer bases on his experience with the dog over approximately 3 weeks.

The results were analysed as a trainers × dogs factorial experiment for which the model is

$$Y = T_i + D_j + e_{ij}$$

where

T_i = the average effect of the ith trainer

D_j = the average score of the jth dog

e_{ij} = an error

The extent to which trainers agreed in their ranking of the dogs is measured by the intra-class correlation coefficient t,
where

$$t = \frac{S_D^2}{S_D^2 + S_e^2}$$

The values obtained for t are given in Table II.

TABLE II

Intra-class correlation coefficients for the trainers' scores

Score	t
Nervousness	0.70
Suspicion	0.62
Concentration	0.65
Willingness	0.0
Distraction	0.45
Dog distraction	0.69
Nose distraction	0.39
Sound shy	0.50
Hearing sensitivity	0.31
Body sensitivity	0.37
Aggression	0.12
Nervous aggression	0.26
Excitability	0.47
Anxiety	0.67

All the scores except N and ND showed significant between-trainer variation. This means that some trainers consistently gave higher scores than others to the 9 dogs.

These average differences between trainers were further examined by performing a 2-way analysis of variance (trainer × year) on the complete data (887 dogs), in which each dog was scored by only one trainer. All the scores showed significant differences between trainers. To make scores from different trainers comparable, the scores were corrected by subtracting the least squares estimates of the effect of each trainer from the raw scores. All further analysis was carried out on these corrected scores. This produces results very similar to analysis carried out "within trainer", e.g. by including the effect of trainers in the model, but was easier to compute.

Correlations between the trainers' scores

The overall correlations between the 17 corrected scores are given in Table III (if correlations are calculated within year, sex, breed and source they are very similar). Several groups of interrelated scores can be distinguished. For instance, the scores measuring fearfulness (N, S, SS, NA and ANX) were all correlated with each other. These groups also appeared in a factor analysis using varimax rotation of principal components. The factor loadings are also shown in Table III. Together, these 5 factors explained 60% of the variance of the original scores. Factor 1 had high loadings for all the distractions. Factor 2 had high loadings for C, W, T and I. It is difficult to see what these 4 scores have in common, and I believe this factor reflects a general assessment by the trainer of a dog's suitability.

Factor 3 was largely "sensitivity" (HS and BS), but also had positive loadings indicating fear (S and ANX) and negative loadings indicating lack of aggression (NA and A) and lack of excitability. It might be considered a submissiveness factor.

The fearfulness scores were split between Factors 4 and 5. Factor 4 had high loadings on N, S, SS and NA, and a lower loading on ANX. Factor 5 had high loadings on ANX and EXC. This suggests that Factor 5 represents an active response to fear as opposed to Factor 4 which is simply fearfulness.

In addition to these high factor loadings which define the factors, there were smaller loadings which indicate correlations between groups of scores. For instance, highly distracted dogs tended to be fearful, excitable and to lack concentration. Highly sensitive, non-aggressive dogs tended not to be dog-distracted, no doubt because all these scores were assessed from the dogs' responses to training to ignore other dogs.

The correlations do not adequately demonstrate the importance of fearfulness in causing aggression. Dogs can be classified as either fearful ($N > 0$ or $NA > 0$) or not fearful ($N = NA = 0$), and as either aggressive ($NA > 0$ or $A > 0$) or not aggressive ($A = NA = 0$) on the uncorrected scores. This results in the contingency tables shown in Table IV. Fearfulness greatly increased the chance that a dog would behave aggressively.

The relationship of the scores to success

Dogs which qualify as guide-dogs are considered to be successful, while those that are rejected in testing or later in training are unsuccessful.

All the scores except HS and BS were significantly correlated with success, but because of correlations between scores not all are needed to predict success.

The best regression equation for predicting success (Y) was

$$Y = -0.015\,N - 0.005\,S + 0.011\,C - 0.010\,DD - 0.010\,A \quad (R^2 = 0.209)$$

A trainer scores a dog at the same time as he decides whether or not to ac-

TABLE III

Overall correlations amongst the trainers' scores and between the trainers' scores and the factors derived from them[1]

		N	S	C	W	D	DD	CD	ND	SS	HS	BS	A	NA	T	I	ANX	EXC
Nervousness	(N)																	
Suspicion	(S)	37																
Concentration	(C)	−31	−30															
Willingness	(W)	−27	−20	60														
Distraction	(D)	12	29	−41	−14													
Dog Distraction	(DD)	03	15	−20	−09	53												
Cat Distraction	(CD)	−01	10	−14	01	41	51											
Nose Distraction	(ND)	08	15	−26	−16	50	38	35										
Sound Shy	(SS)	44	37	−21	−15	13	07	02	03									
Hearing Sensitivity	(HS)	01	16	08	07	03	−11	−07	−07	09								
Body Sensitivity	(BS)	04	13	10	05	−07	−20	−21	−16	04	53							
Aggression	(A)	04	−08	02	−05	03	28	16	14	02	−22	−23						
Nervous Aggression	(NA)	37	14	−08	−10	01	02	04	02	17	−14	−08	14					
Temperamental Stability	(T)	−29	−27	50	36	−24	−18	−11	−16	−24	11	14	−10	−15				
Initiative	(I)	−23	−20	49	42	−16	0	−01	−08	−14	0	0	−03	−09	37			
Anxiety	(ANX)	34	41	−19	−13	22	08	05	14	36	25	21	01	05	−19	−16		
Excitability	(EXC)	11	22	−20	0	37	36	15	21	10	−10	−17	15	13	−30	01	32	
Factor 1			18	−19		68	67	78	64									21
Factor 2		−31	−20	83	79	−28			−22						59	73		
Factor 3			27				−20				79	76	−53	−26	24		32	−32
Factor 4		74	46							68				71	−22		26	
Factor 5			44			33	32			28				−26	−23		68	75

[1] Factor loadings less than 0.15 have been omitted.

The table gives 100 times the correlation coefficient.

Correlations are based on from 833 to 879 dogs, except for those involving ANX or EXC which are based on from 208 to 250.

TABLE IV

The relationship between fearfulness and aggressiveness

	Not fearful	Fearful
Males		
Not aggressive	305	48
Aggressive	47	36
Females		
Not aggressive	319	90
Aggressive	13	29

cept it for training. Therefore, the scores and rejection in testing are not decided independently. However, some dogs are accepted for training but rejected at a later date. Do the trainers' scores predict which dogs will subsequently be rejected in training and which will be successful? The relationship between rejection in training and the scores was significant, but much weaker than the relationship between the scores and rejection in testing. The best prediction equation for success during training (Y) was

$$Y = -0.019\ \mathrm{N} - 0.006\ \mathrm{S} - 0.009\ \mathrm{D} - 0.013\ \mathrm{A} \qquad (R^2 = 0.037)$$

This equation is similar to the last except that D has replaced DD and C, which are correlated with it.

The effect of breed source and sex

To study the effect of these factors on the trainers' scores and on success, a 4-way (breed × source × sex × year) least squares analysis of variance was carried out. Over the 18 variables examined, there were many significant main effects, but only as many significant interactions as would be expected by chance when so many are examined. Also, there was no logical explanation of those interactions which were significant. Therefore, only main effects are reported.

Breed had a significant effect on only 2 scores (CD and BS). Golden Retrievers were more sensitive and less distracted by cats than Labrador Retrievers. However, the test for breed differences is somewhat insensitive since all but 80 of the dogs are Labrador Retrievers. The lack of large breed differences is similar to the results of Scott and Bielfelt (1976), and in part due to the fact that only breeds considered suitable are used.

Source had a significant effect on success rate — A dogs averaged 47%, P dogs 33% and D dogs 19%. It also had a significant effect on 7 scores indicating fear, distraction and aggression (N, S, D, DD, ND, A, NA). In each case, A dogs were superior to P dogs which were superior to D dogs.

Females were more suspicious, excitable and sensitive, less aggressive and

less dog distracted but more nose distracted than males. The superiority of the males on the C, T and I scores reflects their lower level of fearfulness, since these 3 scores correlated negatively with fearfulness.

Year significantly affected 12 of the 17 scores. The pattern is not one of steady improvement but of irregular fluctuations. Nor is there any evidence that the A dogs have steadily increased their superiority over the P dogs.

Heritabilities and genetic correlation

The trainers' scores of 249 association-bred Labrador Retrievers tested between 1970 and 1976 were analysed by the same method as that described in Goddard and Beilharz (1982). The model includes year, sex, sire, dam and residual between-litter variation not explained by the other sources, as well as within-litter error. The scores A, NA, T, I, CD, ANX and EXC were not included in the analysis because there were some missing values for these scores.

From the results of the analysis of variance, the variance components due to the random sources of variations were calculated. These are expressed as a percentage of the total random variation in Table V. Variation between dams was typically less rather than greater than variation between sires. This could occur if the maternal environment effect of dams was negatively correlated with their genetic effect on the offspring; or it could be simply a chance deviation from equal variation due to sires and dams.

Heritability estimates were calculated from the formula

$$h^2 = \frac{2(\sigma^2_{\text{sire}} + \sigma^2_{\text{dam}})}{\sigma^2_{\text{sire}} + \sigma^2_{\text{dam}} + \sigma^2_{\text{residual}} + \sigma^2_{\text{within}}}$$

TABLE V

Variance components and heritabilities for the trainers' scores[1]

Trait	Source of variation				
	Sire	Dam	Residual litter	Within litter	Pooled h^2
Nervousness	19*	10	3	68	58
Suspicion	17	−12	22***	73	10
Concentration	14	0	7	79	28
Willingness	10	0	2	87	22
Distraction	5	−1	17*	79	8
Dog distraction	6	7	7	79	27
Nose distraction	−21	−11	42***	89	0
Sound shy	8	−1	5	88	14
Hearing sensitivity	0	0	9	91	0
Body sensitivity	9	8	3	80	33

[1] All figures are percentages of total variation.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

(In these data this formula results in lower estimates than that based on variation between sires only.)

Nervousness had the highest heritability and was the only trait for which the sire-effect was significant. Except for ND and HS, the other heritability estimates were positive, the highest being BS, C and DD.

Genetic correlations were calculated from pooled sire and dam covariance components (Table VI). The overall pattern of correlations is similar to the pattern of phenotypic correlations in Table III. The undesirable traits correlated positively with other undesirable traits and negatively with desirable traits. The fearfulness traits (N, S, SS) correlated with each other, as did the distraction traits (D, DD).

TABLE VI

Genetic correlations among the trainers' scores[1]

		Success	N	S	C	W	D	DD	SS
Nervousness	(N)	−54							
Suspicion	(S)	−67	53						
Concentration	(C)	+58	−01	−31					
Willingness	(W)	+42	−57	−20	67				
Distraction	(D)	−72	17	140	−91	−50			
Dog distraction	(DD)	−61	11	63	−47	−41	98		
Sound shy	(SS)	−58	89	47	33	−78	58	28	
Body sensitivity	(BS)	+30	72	51	−29	−74	−06	−21	59

[1] The table gives 100 times the correlation coefficient.

Performance of dogs when fully trained

The scores which have been discussed so far are based on the trainers' assessment of the suitability of each dog for guide-dog work. This determines whether a dog is accepted or rejected, and the prime aim of a breeding programme would be to increase the proportion of dogs that are accepted and which qualify as guide-dogs. However, a secondary aim is to improve the performance of dogs that do qualify. To do this, it is necessary to have some assessment of the performance of dogs when fully trained. Therefore, a system in which the trainers rate their dogs on 12 aspects of their work, just before the dogs are given to visually handicapped clients, was begun in 1975. To date, 87 dogs have been rated on this system.

In an attempt to avoid the variation between trainers which occurs in the ordinary trainers scores, a simple 3-point scale was used: (1) satisfactory; (2) good; (3) very good. However, 7 of the 12 scores showed significant between-trainer variation.

Within-trainer correlations amongst the 12 scores were calculated. They are almost entirely positive, suggesting that dogs which are good in one re-

spect tend to be good in all others. However, it could be that a trainer's overall impression of a dog influenced his rating on many different scores.

A 3-way (trainer × sex × source) analysis of variance was performed on these 12 scores. The effect of source was not significant on any score and the effect of sex was significant only for "learns quickly" — females learned more quickly than males.

Some of the ordinary trainers' scores are intended to predict how well a dog will perform during training and therefore should correlate with these ratings of fully-trained performance. To test this hypothesis, the correlations between the corrected trainers' scores and the fully-trained scores were calculated.

Overall the correlations were low but there were more significant correlations than would be expected purely by chance.In general, the undesirable traits — fearfulness and distraction scores — correlated negatively with later performance, whereas concentration correlated positively. In particular, the fearfulness scores (N, S, SS, NA) correlated with performance in fear-provoking situations (stressful situations and strange places). DD and CD correlated negatively with ease of control, as did EXC. BS correlated positively with ease of control. Thus the ordinary trainers' scores had some predictive ability for behaviour similar to that on which they are based. On the other hand, the trainers' scores failed to predict performance on specific learned tasks such as right-shoulder work. Some people believe that a low level of aggression helps a dog in tasks such as pushing through crowds, but there is no evidence for this in these correlations.

DISCUSSION

The extent of agreement between trainers

The scores varied greatly in their repeatability between trainers (Table II). In general, the scores which refer to well-defined responses (e.g. N, S, ANX, DD) had higher repeatabilities than scores for which the response measured is less well defined (W, BS, HS). The aggression scores A and NA had low repeatabilities, in part because little evidence of aggression was seen from any of the dogs.

These intra-class correlation coefficients should not be taken as a direct estimate of the reliability of the normal trainers' scores analysed in the rest of this paper. Some traits can only be evaluated after the trainer has worked with the dogs for some time. For this reason, initiative was not scored in this experiment, and this fact may also contribute to the lack of repeatability found for willingness. On the other hand, these repeatabilities do not take into account variation from day to day in the behaviour of a dog or the effect of different handlers on the dogs' behaviour.

Humphrey and Warner (1934) stated that a dog may be willing for one trainer but not for another. However, they attributed this to differences in

the relationships between dog and trainer rather than to differences in scoring the dog's behaviour.

Correlations between the trainers' scores

Factors 1, 3, 4 and 5 indicate that measurements of similar kinds of behaviour tend to be positively correlated. Positive correlations between different measures of fearfulness in dogs have also been reported by Royce (1955), Brace (1961), Plutchik (1971) and Reuterwall and Ryman (1973). Royce's (1955) factor analysis identified 2 factors indicating fearfulness. One factor involved suppression of physical activity together with a high level of autonomic activity. The second factor indicated a high level of physical activity. These factors are similar to our Factors 4 and 5. However, the distinction between Factor 4 and Factor 5 may be partially an artifact, due to the trainers' tendency to score behaviour as being either suspicion or nervousness or anxiety.

Other correlations may also be influenced by this type of artifact. For instance, if trainers believe that certain behaviour indicates both sensitivity and lack of aggression, these 2 scores will be correlated. The difficulty arises from the lack of precision with which the behaviour measured by each score is defined.

If the group of scores N, S, D and ND are considered, it will be seen that the correlations decrease as one moves away from the diagonal. This indicates that these scores can be arranged in this order along some dimension. N indicates withdrawal, S approach—withdrawal conflict, and D and ND approach behaviour. So the dimension seems to be measuring approach vs. withdrawal. D was correlated with N and S but ND was not. This means that fearful dogs showed a high level of visual and auditory investigation but not olfactory investigation. Increased use of long-range senses when afraid is adaptively appropriate.

Both positive and negative correlations between fearfulness and aggressiveness have been reported (Lagerspetz, 1964; Whitney, 1970; McClearn and DeFries, 1973). Archer (1976) pointed out that noxious stimuli cause both fear and aggression. He suggested that fear dominates at high levels of noxiousness, while aggression dominates at low levels of noxiousness or when escape is not possible. According to this theory, dogs which perceive many stimuli as noxious will be prone to both fearful and aggressive behaviour, as indeed many aggressive dogs were.

Overall, if a dog was good in one score it tended to be good in all other scores. There were no cases of favourable traits being negatively correlated. This fact should make simultaneous improvement in many cases possible.

The relationship of the scores to success

The results agree with the puppy-walking scheme records analysed by Goddard and Beilharz (1982) that the main reasons for rejection were fear-

fulness and distraction, especially dog distraction, and aggression. The results contrast with those reported by Scott and Bielfelt (1976) for the San Rafael Guide-Dog Centre. These authors also found that the most important reasons for rejection were fearfulness, distraction and aggressiveness, but their trainers' scores did not measure these behaviours and so were poorly correlated with rejection. The same traits are also the major reasons for rejection of potential guide-dogs in the U.K. (Guide Dogs for the Blind, 1977). As pointed out by Scott and Bielfelt (1976), dogs fail not through lack of intelligence but because undesirable emotional responses interfere with the correct response.

The effect of source and sex

The difference between A and P dogs is probably genetic, whereas the difference between P and D dogs is probably environmental, due to the effect of rearing on the puppy-walking scheme. There is no information on what aspects of the dogs' experience on the puppy-walking scheme caused the superiority of the P dogs, but the training which they receive and the exposure to a variety of places are obvious possibilities. The sex differences are small considering the strong preference that trainers in the past expressed for females. Either such a preference was mistaken or these scores do not measure the behaviours in which females are greatly superior. Reuterwall and Ryman (1973) also found females to be more fearful and less aggressive than males.

The absence of significant source × sex interactions is surprising. Since D males were entire while A and P males were castrated, one would have expected a greater sex difference in the D dogs than in the A and P dogs. This did not occur, which suggests that the effect of castration on these scores is not great.

In drawing conclusions from these results, 2 sources of error must be kept in mind. Firstly, some dogs of each source were rejected before they were scored by the trainers. Since more P dogs than A dogs were rejected on the puppy-walking scheme, this culling could not have caused the superiority of the A dogs on the trainers' scores. There is no record of the number of D dogs which were rejected before scoring, but many were. However, the culling of the D dogs was probably less accurate than that of the A and P dogs, since it was based on only one observation. This fact may cause the inferiority of the D dogs to be exaggerated. However, it is possible that entire D males showing undesirable male behaviour (e.g. frequent urination) may have been heavily culled, and this could have caused the absence of source × sex interactions discussed above. Secondly, the trainers sometimes know the source of the dogs they are testing and they always know its breed and sex. Therefore, biases on the part of the trainers could have affected the results.

The superiority of the A dogs over the P dogs indicates that the breeding programme has been successful, but the advantage of the A dogs has not con-

tinued to increase over the years. This is probably due to continued introduction of new dogs into the A line. Goddard and Beilharz (1982) came to the same conclusion with respect to overall success rate.

Heritabilities and genetic correlations

The fact that some culls were rejected on the puppy-walking scheme before they could be tested by the trainers could have caused these heritability estimates to be biased. If the worst dogs in each litter were rejected prior to testing, this would reduce between-litter variation and result in the heritability values being under-estimates. On the other hand, if there was really no between-litter variation but some litters were culled more highly than others on the puppy-walking scheme, this would create differences between litters in the dogs remaining and hence over-estimate heritabilities. If this had occurred, it would cause a negative correlation between the proportion of a litter rejected on the puppy-walking scheme and the proportion rejected by the trainers. In fact, the correlation was 0.13, so this source of bias is probably not important.

Genetic variation, and/or successful selection experiments for fearfulness, have been reported by Broadhurst (1960) in rats. De Fries et al. (1974) in mice, Belyaeu and Trut (1975) in foxes, and Scott and Fuller (1965) and Murphree (1973) in dogs. In contrast, Reuterwall and Ryman (1973) estimated the heritability of fearfulness scores in dogs to be low (about 0.1). This may have been due to the fact that the sires used to produce their dogs were a highly selected group, or it may be that their scores were rather unreliable, since they appear to have been based on one observation only. Humphrey and Warner (1934) combined fearfulness and sensitivity into 2 scores (hand-shy and sound-shy), which they claimed to be highly heritable.

The main difference between the genetic correlations and the phenotypic correlations is that the genetic correlations between BS and the fearfulness traits (N, S, SS) were positive and high, whereas the phenotypic correlations were near zero. This difference could be due to the scoring system. A trainer may decide that a dog is not really nervous, just over-sensitive, when, genetically at least, these 2 traits are very similar.

Both BS and the distractions were positively correlated with the fearfulness traits, yet they were negatively correlated with each other. This suggests that BS consists of 2 component traits — "fear of being touched" which is correlated with fearfulness, and "rapid response to correction" which is negatively correlated with distractions. If these 2 postulated component traits are highly correlated, one would expect negative correlations between fearfulness and the distractions. In fact, the correlations in these data were positive, giving support to the suggestion that the trainers really did combine 2 separate component traits into their BS score.

Performance of dogs when fully trained

There are several possible reasons why sex and source had so little effect on these scores. It may be that the culling that had occurred prior to dogs being fully-trained resulted in only dogs of a uniformly high standard remaining. Alternatively, it is possible that the scores were not sufficiently reliable to detect the effects of sex and source, or sex and source may really have very little effect on these behaviours.

Evaluation of the scoring system

All the results of this paper depend on the ability of the trainers to assess the traits which are important in a potential guide-dog. Since they have years of experience in doing this, the results are probably not very misleading. Nevertheless, for scientific analysis there are some inadequacies in their scoring methods. This was demonstrated firstly by the low between-trainer reliabilities of some of the scores. Part of the problem is that the scores are based on the trainers' interpretation of the dog's behaviour, rather than on a description of well-defined classes of behaviour. This fact can cause 2 types of artifacts in the correlations among scores. Firstly, poorly defined classes of behaviour tend to merge with one another, so that the same behaviour affects many supposedly independent scores. This may well explain the correlations among C, T, W and I. Also, when behaviour is not recorded immediately, the trainer may simply remember the dogs as good or bad and base several scores on this general impression. This may have caused the uniform positive correlations among the scores of fully-trained performance. Secondly, the interpretative nature of the scores leads the trainer to decide that a dog is really suspicious not nervous, or nervous not over-sensitive, when in fact these scores are all measuring similar highly correlated behaviours. This causes the correlations between traits to be reduced, and may have caused the low heritability estimate for suspicion.

Another problem occurs when one score measures 2 types of behaviour which are themselves almost uncorrelated. BS combines rapid response to correction and handling, which is desirable, with fear of handling, which is undesirable. If these 2 traits were highly correlated, it would have important implications regarding selection for low fearfulness, but because the 2 traits are represented by only one score, it is impossible to estimate the correlation between them.

CONCLUSION

The most common reasons why the trainers rejected dogs were fearfulness, distraction, especially dog distraction, and aggression. The selection programme seems to have been successful in improving all of these traits. Fearfulness was moderately heritable ($h^2 = 0.5$) and should respond to future selection. The

heritability estimate for dog distraction was 0.27 but non-significant. There were no unfavourable correlations between traits, so it should be possible to make an overall improvement in the dogs by future selection. Selection against fearfulness would possibly also decrease distraction and aggression and slightly increase concentration, willingness and initiative.

Although insufficient data were available for genetic analysis of the performance of fully-trained dogs, the phenotypic correlations suggest that decreasing fearfulness would also improve final performance.

The poor success-rate of donated adults compared to donated puppies shows the value of the puppy-walking scheme. However, a full evaluation of its usefulness would have to include consideration of the costs of rejecting dogs at the various stages of rearing and training.

ACKNOWLEDGEMENT

This work was supported by a grant from the Uncle Bens Company in Australia and overseas. The authors are most grateful for the cooperation of the Royal Guide Dogs for the Blind Association of Australia.

REFERENCES

Archer, J., 1976. The organization of aggression and fear in vertebrates. In: P.P.G. Bateson and P.H. Klopfer (Editors), Perspectives in Ethology, Vol. 2. Plenum Press, New York.

Belyaeu, D.R. and Trut, L.N., 1975. Some genetic and endocrine effects of selection for domestication in silver foxes. In: M.W. Fox (Editor), The Wild Canids. Van Nostrand Reinhold, New York, pp. 416—426.

Brace, C.L., 1961. Physique, physiology and behaviour: an attempt to analyse a part of their roles in the canine biogram. Ph.D. Thesis, Harvard University. Quoted by Scott and Fuller, 1965.

Broadhurst, P.L., 1960. Experiments in psychogenetics: applications of biometrical genetics to the inheritance of behaviour. In: H.J. Eysenck (Editor), Experiments in Personality, Vol. 1. Psychogenetics and Psychopharmacology, Routledge, London.

DeFries, J.C., Hegmann, J.P. and Halcomb, R.A., 1974. Response to 20 generations of selection for open field activity in mice. Behav. Biol., 11: 481—495.

Goddard, M.E. and Beilharz, R.G., 1982. Genetic and environmental factors affecting the suitability of dogs as guide-dogs for the blind. Theor. Appl. Genet., in press.

Guide Dogs for the Blind, 1977. Annual Training Report 1976/77. The Guide Dogs for the Blind Association, Windsor, Berkshire.

Humphrey, E. and Warner, L., 1934. Working Dogs. Johns Hopkins Press, Baltimore.

Lagerspetz, K.M.J., 1964. Studies on the aggressive behaviour of mice. Ann. Acad. Sci. Fenn., Ser. B, 131: 1—131.

McClearn, G.E. and DeFries, J.C., 1973. Genetics and mouse aggression. In: J.F. Krutson (Editor), The Control of Aggression: Implications from Basic Research. Aldene, Chicago.

Murphree, O.D., 1973. Inheritance of human aversion and inactivity in two strains of pointer dog. Biol. Psychiatry, 7: 23—29.

Pfaffenberger, C.J., 1963. The New Knowledge of Dog Behaviour. Howell, New York.

Plutchik, R., 1971. Individual and breed differences in approach and withdrawal in dogs. Behaviour, 40: 302—311.

Reuterwall, C. and Ryman, R., 1973. An estimate of the magnitude of additive genetic variation of some mental characters in Alsatian dogs. Hereditas, 73: 277—284.

Royce, J.R., 1955. A factorial study of emotionality in the dog. Psychol. Monogr., 69: 1–27.

Scott, J.P. and Beilfelt, S.W., 1976. Analysis of the puppy testing program. In: C.J. Pfaffenberger, J.P. Scott, J.L. Fuller, B.E. Ginsburg and S.W. Bielfelt. Guide Dogs for the Blind: Their Selection, Development and Training. Elsevier, Amsterdam.

Scott, J.P. and Fuller, J.L., 1965. Genetics and social behaviour of the dog. University of Chicago Press, Chicago.

Whitney, G.D., 1970. Aggression and open field behaviour in male mice. Psychon. Sci., 18: 167—168.