



**University of
Nottingham**

UK | CHINA | MALAYSIA

**Genetic influences on puppy and juvenile guide dog
behaviour**

By

Rebecca Hunt

20217603

Thesis submitted to the University of Nottingham for the Degree of Doctor
of Philosophy

February 2025

University of Nottingham
School of Veterinary Medicine and Science
Sutton Bonington Campus
Leicestershire LE12 5RD

Abstract

The influence of genetics on behaviour in dogs has been under constant debate and is of particular importance to assistance dog organisations, where breeding suitable dogs for the working role is important to maximise resources and efficiencies and improve dog welfare. The aims of this thesis were to determine the genetic influences on guide dog behaviour through pedigree and genome-based heritability estimates for data captured through questionnaires and a behavioural observation assessment. Behavioural questionnaires were collected in a cohort of juvenile guide dog puppies at five months (5M, n=551), eight months (8M, n=2317) and 12 months (12M, n=1537) of age using a previously validated questionnaire, the refined-Puppy Walker Questionnaire (r-PWQ). Due to a difference in component (group of questions) structure for this population, the questionnaire was adjusted. A pedigree-based heritability analysis was undertaken, with the 5M Anxiety component indicating the highest heritability ($h^2=0.29$). Heritabilities at 8M and 12M were below 0.1 for all behavioural components. Using the same data, a genomic-based heritability analysis was undertaken for 96 dogs at 8M and 12M. Genomic-based heritabilities were higher than pedigree-based estimates and following a genome wide association study (GWAS), Excitement behaviours at 12M and Separation behaviours at 8M indicated single nucleotide polymorphisms (SNPs) approaching significance. Links were found between these behaviours in dogs and genes associated with obsessive/depressive disorders in humans (attention deficit hyperactivity disorder, alcoholism, depression and schizophrenia).

Behavioural assessments comprising of six published tasks were undertaken in a cohort of 399 puppies around eight weeks of age. Twenty-two of the 79 behavioural measures demonstrated heritabilities above 0.4, with the highest heritabilities found for the Novel Object and Surprising Events tasks. A principal component analysis (PCA) extracted six behavioural components, demonstrating correlations to questionnaire components, supporting use of the questionnaire at Guide Dogs UK

(GDUK). Evaluation of inter and intra-reliability of behaviours observed during the assessment indicated differences between assessors, highlighting the importance of consistency and reliability in behavioural assessments. Following these findings, a pilot study utilising machine learning (ML) was undertaken, comparing computer generated and human generated results. It is suggested that future studies could utilise ML methods for behavioural analysis in puppies.

Behaviours associated with anxiety/fear were shown to be consistently heritable in this population through pedigree-based and genomic-based methods, suggesting that some of the variation in the phenotype displayed can be attributed to genetic influences. Based on the findings from the present study, GDUK could select away from these behaviours during breeding should it be possible to account for the behaviours when planning matings between sires and dams. The present study also highlighted environmental and puppy raiser influences on behaviour; GDUK should continue to invest in supporting volunteers through puppy raising due to the large potential to positively impact future behaviour.

The use of terminology when describing and discussing behaviour has been under debate by behavioural researchers, with the terms ‘temperament’ and ‘personality’ often used interchangeably and with few stating definitions for the terms. The present study discusses the two terms, and a conclusion made that the behavioural data analysed during the thesis is indicative of ‘temperament’, with reasonings and evidence discussed.

Acknowledgements

This thesis would not have been possible without the dedicated support of my supervisors, Professor Gary England and Dr Sarah Blott. Thank you for your time, patience, expertise and kindness over the past six years. I have learnt a great deal working with you both.

Thanks to Guide Dogs for supporting this project and to the colleagues that have provided me with assistance and guidance during the PhD. Thanks to Guide Dogs puppy raisers for completing behavioural questionnaires and to the wonderful puppies I worked with during behavioural assessments.

I am grateful to have worked with some other talented researchers during this thesis, including Dr Emily Bray and Dr Robert Hyde. Your knowledge enabled me to learn a great deal.

Thanks to my family, my partner Chris, and my friends for your continued love and support - I know I haven't been the best company at times!

Lastly, but by no means least, my wonderful Golden Girls (Charlie, Bella and Crumble). I am lucky to have grown up with such special dogs, leading me to work at Guide Dogs and to undertake this thesis.



Ethical approval

This study received the following ethical approval by the School of Veterinary Medicine and Science, Ethical Review Committee:

- Ethical approval given 30/08/2019 for collection of questionnaires (reference 2798 190814)
- Ethical approval given 02/11/2020 for collection of genotype samples (reference 3248 200108)
- Ethical approval given 20/12/2021 for behavioural observation assessments (reference 3501 211130)

Statistical analysis acknowledgement

It is acknowledged, with thanks, that the pedigree file used for heritability analyses in Chapters 3 and 4 was prepared and recoded by Dr Tom Lewis. Heritability models were undertaken by RH. The ethogram used in Chapter 4 was provided by Dr Emily Bray. The machine learning models and analysis in Chapter 5 (Part 2) were undertaken by Dr Robert Hyde. RH supported with analysis by cropping videos clips, manually placing markers on subjects for model training and timing interactions as a coder.

Publications arising from thesis

Conference presentations

- “Can we predict guide dog success by measuring puppy behaviour as early as eight months of age?” – International Guide Dog Federation Conference (virtual) 28/09/2021.
- “Improving behaviour monitoring within a working dog programme” – International Working Dog Conference (virtual) 10/12/202.
- “Puppy Cognition” – ASAB Animal Cognition and Animal Welfare Workshop, Devon UK, 20-22/09/2022.
- “Improving Behaviour Monitoring Within a Large Guide Dog Organisation Whilst Meeting the Needs of Breeding, Puppy Raising and Training Departments” – International Guide Dog Federation Conference, Vancouver Canada, 27-30/04/2023.

Abbreviations

5M – 5 month behavioural questionnaire

8M – 8 month behavioural questionnaire

12M – 12 month behavioural questionnaire

C-BARQ – Canine Behavioral Assessment & Research Questionnaire

GDIR – Guide Dogs Interactive Replacement GDIR.

GDNC – Guide Dogs National Centre

GDUK – Guide Dogs UK

GR – Golden Retriever

GSD – German Shepherd Dog

GWAS – genome wide association study

Lab – Labrador

ML – machine learning

PCA – principal component analysis

PBD – Potential Breeding Dogs

PDA – Puppy Development Advisors

PR – puppy raisers

PWQ – puppy walker questionnaire

REML - restricted maximum likelihood

r-PWQ – refined puppy walker questionnaire

r-PWQ2 – refined puppy walker questionnaire version 2

SNP – single nucleotide polymorphism

Table of contents

Abstract.....	i
Acknowledgements	iii
Ethical approval	iv
Statistical analysis acknowledgement.....	iv
Publications arising from thesis	v
Abbreviations.....	vi
Table of contents	vii
Figures.....	xii
Tables.....	xiv
Chapter 1. Literature review and background reading	1
1.1 Dogs and assistance dogs	1
1.1.1 Assistance dogs.....	2
1.2 Behaviour, temperament and personality.....	3
1.2.1 Behaviour, temperament and personality in dogs	6
1.2.2 Methods to monitor behaviour and temperament in dogs ...	12
1.2.3 Guide dog behaviour monitoring	18
1.3 Breeding selection	20
1.3.1 Phenotypic selection	21
1.3.2 Genomic selection	22
1.3.3 Heritability	25
1.3.4 Behavioural genetics.....	29
1.3.5 Guide dog breeding programmes	32
1.4 Conclusion	36
Chapter 2. Refining a puppy raiser questionnaire from the r-PWQ to r-PWQ2.....	38
2.1 Introduction	38
2.2 Aims and Objectives	40
2.2.1 Research question	40
2.2.2 Aim.....	41
2.2.3 Objectives	41
2.3 Materials and methods.....	41
2.3.1 Study design	41
2.3.2 Study setting	43
2.3.3 Questionnaires.....	46

2.3.4	Variables and data sources	47
2.3.5	Bias and study size	47
2.3.6	Quantitative variables and statistical methods.	48
2.4	Results	53
2.4.1	Participants	53
2.4.2	Principal component analyses (PCA).....	57
2.4.3	Refined Puppy Walker Questionnaire version 2 (r-PWQ2) .	67
2.4.4	Comparing the r-PWQ and r-PWQ2.....	72
2.4.5	Additional factors	76
2.5	Discussion.....	86
2.6	Conclusion	93
Chapter 3. Heritability analysis of a behavioural questionnaire (the r-PWQ2) for guide dog puppies at five, eight and 12 months of age ...		94
3.1	Introduction	94
3.2	Aims and Objectives	98
3.2.1	Research question	98
3.2.2	Aim.....	98
3.2.3	Objectives	98
3.3	Materials and methods.....	98
3.3.1	Study design and setting.....	98
3.3.2	Study animals	99
3.3.3	Variables and data sources	99
3.3.4	Bias and study size	101
3.3.5	Quantitative variables and statistical methods	101
3.4	Results	103
3.4.1	5M heritability.....	104
3.4.2	8M heritability.....	105
3.4.3	12M heritability.....	107
3.4.4	Summary of heritability results	109
3.5	Discussion.....	109
3.6	Conclusion	119
Chapter 4. Heritability of the behavioural responses of guide dog puppies to a series of problem-solving tasks		120
4.1	Introduction	120
4.2	Aims and Objectives	123
4.2.1	Research question	123
4.2.2	Aim.....	123
4.2.3	Objectives	124

4.3	Materials and methods.....	124
4.3.1	Study design	124
4.3.2	Study setting	124
4.3.3	Study animals	127
4.3.4	Assessment protocol.....	128
4.3.5	Variables and data sources	137
4.3.6	Bias and study size	147
4.3.7	Quantitative variables and statistical methods	148
4.4	Results	150
4.4.1	Assessment data	150
4.4.2	Behavioural analysis	152
4.4.3	Breed, sex and age effects	155
4.4.4	Heritability of behavioural measures	157
4.4.5	Linking 5M r-PWQ2 scores to heritable behavioural responses.....	160
4.5	Discussion.....	165
4.6	Conclusion	173
Chapter 5. Reliability of behavioural measures collected during a puppy assessment by human coding and machine learning methods.....		174
5.1	Introduction	174
5.2	Aims and Objectives	178
5.2.1	Research question	178
5.2.2	Aim.....	178
5.2.3	Objectives	178
5.3	Materials and methods.....	179
5.3.1	Study design	179
5.3.2	Study setting	179
5.3.3	Study animals	179
5.3.4	Variables and data sources	181
5.3.5	Bias and study size	183
5.3.6	Quantitative variables and statistical methods	184
5.4	Results	188
5.4.1	Part 1 - Inter and intra-rater reliability.....	188
5.4.2	Part 2 – Machine learning pilot	190
5.5	Discussion.....	194
5.6	Conclusion	199

Chapter 6. An exploratory genome-wide analysis of r-PWQ2 components at eight and 12 months	201
6.1 Introduction	201
6.2 Aims and Objectives	205
6.2.1 Research question	205
6.2.2 Aim.....	205
6.2.3 Objectives	205
6.3 Materials and methods.....	206
6.3.1 Study design	206
6.3.2 Study setting	206
6.3.3 Study animals	206
6.3.4 Variables and data sources	208
6.3.5 Bias and study size	209
6.3.6 Quantitative variables and statistical methods	210
6.4 Results	216
6.4.1 r-PWQ2 data.....	216
6.4.2 Sample quality control checks	219
6.4.3 12M REML heritability analysis	222
6.4.4 12M GWAS.....	223
6.4.5 8M REML heritability analysis	233
6.4.6 8M GWAS	235
6.4.7 GWAS results summary.....	244
6.5 Discussion.....	245
6.5.1 Excitement associated behaviours.....	249
6.5.2 Separation associated behaviours	251
6.5.3 Future research.....	253
6.6 Conclusion	255
Chapter 7. General discussion and conclusions	256
7.1 Previous studies undertaken at GDUK.....	257
7.2 Reliability of behaviour data	260
7.3 Pedigree-based and genome based heritabilities.	263
7.4 Anxiety-related behaviours.....	268
7.5 Terminology conclusions.....	273
7.5.1 Updated behaviour schematic	277
7.6 Selective breeding for behaviour.....	281
7.7 Wider implications	283
7.8 Study limitations	286
7.9 Future work at GDUK.....	290

7.10 Conclusion	292
Chapter 8. Appendices	294
8.1 r-PWQ email and survey introduction.....	294
8.2 r-PWQ questionnaire.....	296
8.3 PCA extraction results.....	299
8.3.1 5M PCA extraction results	299
8.3.2 8M PCA extraction results	303
8.3.3 12M PCA extraction results	307
8.4 r-PWQ2 component distribution	311
8.4.1 5M component r-PWQ2 distributions	311
8.4.2 8M component r-PWQ2 distributions	313
8.4.3 12M component r-PWQ2 distributions	315
8.5 r-PWQ2 factor averages	317
8.5.1 5M factor averages	317
8.5.2 8M factor averages	320
8.5.3 12M factor averages	323
8.6 log likelihood comparisons for h^2 models	326
8.6.1 5M heritability models	326
8.6.2 8M heritability models	327
8.6.3 12M heritability models	328
8.7 r-PWQ2 fixed factor effects	329
8.8 Behavioural assessment data sheet	331
8.9 Equipment used for behavioural assessment	333
8.10 Behavioural measures summary.....	334
8.11 Behavioural measures heritability	339
8.12 Behavioural assessment reliability models.....	343
8.13 Genotype sample paperwork	345
References	348

Figures

Figure 2.1 Development of the r-PWQ2.....	43
Figure 2.2 Flow diagram of GDUK journey	45
Figure 3.1 Breed calculation for 0.5 Labrador.....	100
Figure 3.2 Breed calculation for 0.25 Labrador.....	100
Figure 4.1 Behavioural assessment room	125
Figure 4.2 Testing mat used for behavioural assessments.....	126
Figure 4.3 Diagram of assessment room setup	126
Figure 4.4 Mock up imagery of Unsolvable.....	129
Figure 4.5 Mock up imagery of Human Interest	130
Figure 4.6 Mock up imagery of Odour Discrimination.	131
Figure 4.7 Mock up imagery of Novel Object.	132
Figure 4.8 Mock up imagery of Cylinder	134
Figure 4.9 Mock up imagery of Surprising Events (bin bag).	135
Figure 4.10 Mock up imagery of Surprising Events (umbrella).	136
Figure 4.11 Mock up imagery of Surprising Events (metal sheet).....	136
Figure 4.12 Red square to define Novel Object proximity to cat.....	141
Figure 4.13 Scoring protocol for Novel Object reapproach zones.....	141
Figure 4.14 Red square to define Surprising Events proximity to object	145
Figure 4.15 Scoring protocol for Surprising Events reapproach zones	145
Figure 4.16 Puppy assessments undertaken by month	151
Figure 5.1 Coding schedule for inter and intra-rater reliability	182
Figure 5.2 Skeleton used for machine learning	185
Figure 5.3 Data management pipeline for machine learning	187
Figure 5.4 Behaviour model predictions for ML pilot.....	193
Figure 6.1 12M r-PWQ2 component score distribution	217
Figure 6.2 8M r-PWQ2 component scores distribution	218
Figure 6.3 Heterozygosity estimates for genotypes samples.....	219
Figure 6.4 Breed dimensions outputs for pure and crossbreeds.....	221
Figure 6.5 QQ plot for C6 at 12M.....	225
Figure 6.6 Manhattan plot for C6 at 12M.	225
Figure 6.7 QQ plot for C7 at 12M.....	227
Figure 6.8 Manhattan plot for C7 at 12M.	227

Figure 6.9 QQ plot for C8 at 12M.....	232
Figure 6.10 Manhattan plot for C8 at 12M.	233
Figure 6.11 QQ plot for C3 at 8M.....	236
Figure 6.12 Manhattan plot for C3 at 8M.	237
Figure 6.13 QQ plot for C5 at 8M.....	241
Figure 6.14 Manhattan plot for C5 at 8M.	242
Figure 6.15 QQ plot for C8 at 8M.....	243
Figure 6.16 Manhattan plot for C8 at 8M.	244
Figure 7.1 Updated schematic for temperament and behaviour.	280
Figure 8.1 r-PWQ2 email template	294
Figure 8.2 r-PWQ2 questionnaire opening page.....	295
Figure 8.3 Genotype study participation email.....	345
Figure 8.4 Genotype study participation request letter	346
Figure 8.5 Genotype sample collection instructions.....	347

Tables

Table 1.1 Temperament traits for dogs.....	9
Table 2.1 Internal reliability of 5M, 8M and 12M r-PWQ traits	48
Table 2.2 Breeds/crossbreeds represented in the r-PWQ2 study population.	54
Table 2.3 Components extracted from 5M PCA.	58
Table 2.4 Internal reliability for 5M PCA	58
Table 2.5 Components extracted from 8M PCA.	59
Table 2.6 Internal reliability for 8M PCA	60
Table 2.7 Components extracted from 12M PCA.	61
Table 2.8 Internal reliability for 12M PCA	61
Table 2.9 Summary of component Cronbach's alpha	62
Table 2.10 r-PWQ2 component structure.	64
Table 2.11 5M r-PWQ2 component scores.....	67
Table 2.12 Sex and breed effects for 5M behavioural components	68
Table 2.13 8M r-PWQ2 component scores.....	68
Table 2.14 Sex and breed effects for 8M behavioural components	69
Table 2.15 12M r-PWQ2 component scores.....	70
Table 2.16 Sex and breed effects for 12M behavioural components	71
Table 2.17 r-PWQ and r-PWQ2 component correlation at 5M.....	72
Table 2.18 r-PWQ and r-PWQ2 component correlation at 8M.....	72
Table 2.19 r-PWQ and r-PWQ2 component correlation at 12M.....	73
Table 2.20 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 5M	74
Table 2.21 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 8M	75
Table 2.22 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 12M	75
Table 2.23 Puppy raiser experience	77
Table 2.24 Additional factor information for r-PWQ2 questionnaires	78
Table 2.25 Additional factor missing information for r-PWQ2 questionnaires	78
Table 2.26 Significant factors for r-PWQ2 components at 5M.	80

Table 2.27 Significant factors for r-PWQ2 components at 8M.	82
Table 2.28 Significant factors for r-PWQ2 components at 12M.	83
Table 2.29 Summary of factor significance	86
Table 3.1 Puppy breeds represented in r-PWQ2 heritability analyses...	103
Table 3.2 5M heritability estimates for r-PWQ2 components.....	104
Table 3.3 8M heritability estimates for r-PWQ2 components.....	106
Table 3.4 12M heritability estimates for r-PWQ2 components.....	107
Table 4.1 Breed proportions for puppies in behavioural assessments...	127
Table 4.2 Novel Object ethogram	139
Table 4.3 Surprising Events ethogram.....	143
Table 4.4 Behavioural assessment heritabilities	158
Table 4.5 Behavioural assessment components	159
Table 4.6 5M r-PWQ2 scores for puppies in behavioural assessments.	160
Table 4.7 5M r-PWQ2 and behavioural measures correlation	162
Table 4.8 5M r-PWQ2 and behavioural assessment component correlation.	164
Table 5.1 Breed and sex of puppies for reliability studies.....	180
Table 5.2 Inter and intra-rater behavioural measures	184
Table 5.3 Inter (R1) and intra-rater (R2) reliability	188
Table 5.4 Agreement between scorers for ML model validation	191
Table 5.5 Agreement between scorers for ML test data set	192
Table 6.1 12M Genomic-based heritability estimates for 12M r-PWQ2 components.	222
Table 6.2 12M Genomic-based and pedigree-based heritability estimates for 12M r-PWQ2 components.	223
Table 6.3 Top 20 significant SNPs for C6 at 12M	224
Table 6.4 Top 20 significant SNPs for C7 at 12M	226
Table 6.5 Associated behavioural phenotypes for C7 at 12M.....	229
Table 6.6 Top 20 significant SNPs for C8 at 12M	231
Table 6.7 8M Genomic-based heritability estimates for 8M r-PWQ2 components.	233
Table 6.8 8M Genomic-based and pedigree-based heritability estimates for 8M r-PWQ2 components.	234

Table 6.9 Top 20 significant SNPs for C3 at 8M	235
Table 6.10 Associated behavioural phenotypes for C3 at 8M.....	238
Table 6.11 Top 20 significant SNPs for C5 at 8M	240
Table 6.12 Top 20 significant SNPs for C8 at 8M	242
Table 7.1 Summary of data, population size and analysis by Chapter ..	287
Table 8.1 r-PWQ 2 question component group, question number (N) and wording	296
Table 8.2 Additional information questions	297
Table 8.3 5M PCA extraction for r-PWQ2 questions.....	299
Table 8.4 Correlation matrix for 5M PCA	301
Table 8.5 8M PCA extraction for r-PWQ2 questions.....	303
Table 8.6 Correlation matrix for 8M PCA	305
Table 8.7 12M PCA extraction for r-PWQ2 questions.....	307
Table 8.8 Correlation matrix for 12M PCA	309
Table 8.9 Influence of sex and breed on 5M r-PWQ2 scores	317
Table 8.10 Influence of additional factors on 5M r-PWQ2 scores.....	318
Table 8.11 Influence of sex and breed on 8M r-PWQ2 scores	320
Table 8.12 Influence of additional factors on 8M r-PWQ2 scores.....	321
Table 8.13 Influence of sex and breed on 12M r-PWQ2 scores	323
Table 8.14 Influence of additional factors on 12M r-PWQ2 scores.....	324
Table 8.15 Model outputs for 5M C4 model 1	326
Table 8.16 Model output for 5M C4 model 2.....	326
Table 8.17 Model outputs for 8M C 5 model 1	327
Table 8.18 Model outputs for 8M C5 model 2	327
Table 8.19 Model outputs for 12M C 6 model 1	328
Table 8.20 Model outputs for 12M C 6 model 2	328
Table 8.21 Fixed factor effects for 5M heritability outputs.....	329
Table 8.22 Fixed factor effects for 8M heritability outputs.....	329
Table 8.23 Fixed factor effects for 12M heritability outputs.....	330
Table 8.24 Behavioural assessment measures	334
Table 8.25 Heritabilities, sex, breed and age effects for behavioural measures.....	339
Table 8.26 Reliability models for behavioural measures.....	343

Chapter 1. Literature review and background reading

1.1 Dogs and assistance dogs

The dog (*Canis familiaris*) is the only member of the Canidae family that is fully domesticated (Serpell, 1995) and it is estimated that human-dog relationships date as far back as 100,000 years ago (Tami and Gallagher, 2009). Dogs were reported as one of the first companion animals (Topál *et al.*, 1997) and today there are an estimated 10.6 million pet dogs in the UK with 28% of UK adults owning a dog (PDSA, 2024). Dogs are also selected and trained for several working roles, and the variety of these roles has expanded over the years. Working dog roles include herding, police work, search and rescue, military work and as assistance dogs for people with a range of disabilities (Coppinger and Coppinger, 2001).

Dogs have one of the greatest within-species morphological diversities amongst mammals. It is thought that there are between 350 and 400 breeds of dogs (Boyko, 2011), with many more crossbreeds emerging as a result of extensive selective breeding. Along with the great difference in morphology comes behavioural differences between breeds. Within dog research, topics such as differences in phenotype (for both health and behaviour) and genetics between and within breeds, dog and owner bond/interactions, undesirable behaviours and reasons for people giving up dogs to shelters are of particular interest. Companion animal behaviour problems can significantly impact animal and owner welfare, and it is important that behavioural issues in dogs are understood and reduced where possible. Dinwoodie *et al.* (2019) reported the prevalence of behaviour problems in a population of 4114 dogs (76% of dogs were from the United States and the remainder from 16 other countries) to be 85%, highlighting the need to understand and improve dog behaviour. In recent years there has been an increase in the promotion of responsible dog ownership (Jarvis, 2018) and more focus on training dogs and improving welfare. When examining the concept of 'responsible dog ownership',

Westgarth *et al.* (2019) claimed the term to be perceived and interpreted differently amongst dog owners and therefore is of limited use as a message to promote behaviour change. However, it could be argued that improved education for owners and the adoption of training and enrichment practices by owners could reduce the prevalence of behavioural problems and is therefore a benefit to dog welfare and public health.

1.1.1 Assistance dogs

Assistance dogs are raised and trained by organisations to perform specific tasks beneficial to humans (Coppinger and Coppinger, 2001; Bassan *et al.*, 2024). In the UK there are 14 assistance dog charities supporting over 7000 people with disabilities by providing specialist trained dogs for individuals with vision impairments, hearing loss, medical conditions, mobility issues and emotional support (ADUK, 2025). Guide dogs are the most common type of assistance dogs (Sachs-Ericsson *et al.*, 2002), being trained to assist individuals with vision impairment within the community (Mizukoshi *et al.*, 2008; Valsecchi *et al.*, 2010). Guide dogs are primarily Labrador, Golden Retriever and German Shepherd Dogs, or crosses of these breeds (Willis, 1989). Guide dogs arguably have the greatest demands amongst assistance dogs, due to the safety of the client being reliant on successful guiding by the dog (Audrestch *et al.*, 2015). Guide dogs are responsible for leading and alerting the client to a range of hazards and obstacles such as overhanging tree branches, cars parked on pavements and silent electric cars, providing complex challenges for the dog to overcome whilst working. Guide dogs are reported to have considerable benefits to individuals with vision impairments including increased confidence, ability to travel independently and access to wider opportunities (Whitmarsh, 2005; McIver *et al.*, 2020; Rodriguez *et al.*, 2020) and demand for the service is growing (Funk and Williams, 2020). It is therefore essential for guide dog organisations to maximise resources and ensure qualification rates are as high as possible due to the high

costs associated with breeding, rearing and training dogs. In 2023, the guide dog service cost Guide Dogs UK £73.9million (Guide Dogs, 2025) and success qualifying as a guide dog is around 50% for dogs born alive (Moxon and Fellows, 2024).

1.2 Behaviour, temperament and personality

In animal behaviour, there are three terms commonly used within the literature: 'behaviour', 'temperament' and 'personality'. Behaviour is generally defined as the observable and quantifiable response of an animal to stimuli or events within the environment (Levitis *et al.*, 2009). The display of behaviour is commonly referred to as a behavioural phenotype and is influenced by the animal's genotype and environmental factors experienced by the individual (Mackenzie *et al.*, 1986; Falconer and Mackay, 1989). However, a common problem in the field of behaviour are the inconsistencies between studies in the traits or behaviours measured, often without definitions provided. Levitis *et al.* (2009) suggested that scientists do not agree on what constitutes 'behaviour', with researchers commonly contradicting published definitions. To build on the work of others, and to expand knowledge on specific behavioural traits, consistency in recording and descriptions is required (Bellamy *et al.*, 2018).

The terms temperament and personality are often used interchangeably (McCrae *et al.*, 2000) and the distinction between the two is not always clear. Shiner and DeYoung (2013) argued that temperament and personality are different ways of describing the same traits within humans. The researchers report temperament to focus on early emerging individual differences and personality to be focused on juvenile development into adulthood. Similarly, Jones and Gosling (2005) undertook a review of temperament and personality in dogs and found little difference in the definitions of the two terms. The authors reported the main difference to be the disciplines of researchers involved, with animal researchers

commonly using the term ‘temperament’ and human researchers using ‘personality’. Both temperament and personality commonly refer to the consistent variation in behaviour between individuals, and both are commonly described as a number of defined traits (Arvidsson *et al.*, 2017). Both are also influenced by genes, experiences and the environment (Shiner and DeYoung, 2013), so to define differences between the two becomes challenging and it is important to consider differences at the phenotypic level.

The meaning of ‘temperament’ is difficult to define (Sforzini *et al.*, 2009). Numerous definitions exist for temperament, but the term is generally defined as innate individual differences in behaviour that have the ability to influence or predispose an individual in the development and display of specific behavioural traits (McCrae *et al.*, 2000; Rothbart *et al.*, 2000; Putnam, 2011). Jones and Gosling (2005) stated temperament “influences an individual’s behaviour and responses to the environment” and that animal researchers do not always define temperament in the same way. Temperament can often be referred to as a construct of personality, but researchers have discussed how using the term personality could be considered anthropomorphic (Mehta and Gosling, 2008), which is why the term personality may be primarily used within human research. However, studies in humans have also used the term temperament, being defined as the inherited, early appearing tendencies that continue throughout life, which serve as the foundation for personality (Goldsmith *et al.*, 1987; Buss, 1995; Jones and Gosling, 2005). This leads to confusion between the two terms, and which is appropriate to use.

There are also many definitions for ‘personality’ within literature. Frequently personality is defined as characteristics that differ between individuals and account for consistent patterns of behavioural responses to stimuli and situations (Jones and Gosling, 2005; Penke *et al.*, 2007; Kubinyi *et al.*, 2009; Briffa and Weiss, 2010; Mirkó *et al.*, 2012; Ilska *et al.*, 2017), a definition which is very similar to ‘temperament’. The term

personality is also used interchangeably between species which adds further misunderstanding to the term. The foundations for understanding personality began in humans and, consequently a large proportion of the literature focusses on humans. Rothbart *et al.* (2000) reported that personality in humans goes beyond temperament as personality includes an individual's view of the self, other people, the world, links between themselves and other concepts and cognitive adaptations to the world. All these factors, whilst they could be present in the animal domain, are factors that researchers would be unable to study and understand accurately. As the Rothbart *et al.* (2000) study was undertaken in humans, findings are not applicable to all other animals. Some researchers, however, discuss that personality is found within a range of species including mammals (Mehta and Gosling, 2008; Réale *et al.*, 2009), birds (Garamszegi *et al.*, 2008; Kurvers *et al.*, 2010) and reptiles (Siviter *et al.*, 2017). Although, it could be argued that these studies could have referred to temperament and reported the same findings. Gosling *et al.* (2003) claim to have compared personality ratings in humans to those of dogs, reporting judgments in animals and humans to be similar and therefore that personality traits can be measured in animals using similar methods to humans. Still, it seems unlikely that it would be possible to compare such distinct species, and the findings of the researchers may be misleading as animal personality is defined and scored by humans. It is arguable that personality between species is not directly comparable, and so studies may be reporting a different understanding of 'personality'.

Harvey (2014) debated the difference between temperament and personality when designing a questionnaire for guide dog puppies, concluding that personality in dogs could be measured. Harvey produced a schematic, proposing interactions between the terms, with personality defined as being influenced by both temperament and environment. This supports the concept that personality is a higher concept than temperament, similar to the conclusions made by Rothbart *et al.* (2000), albeit the latter centred around humans. Harvey (2014) defined

temperament as being able to influence or predispose an individual to the development of personality traits, and thus a dog's personality is based on their temperament and environmental experience. Harvey's choice of term for dogs is personality. Similarly, Mehrkam and Wynne (2014) suggest that personality is due to interactions between temperament and the environment. This further highlights the difference in the use of terminology between researchers, and the importance of researchers defining temperament or personality in behavioural studies.

As discussed above, the terms temperament and personality are often used interchangeably, and arguably report similar behavioural tendencies in individuals over a multitude of environments. Both terms refer to specific traits, both are influenced by genetics and environment and researchers commonly report either temperament or personality dependent on the individual's preferences, experiences and field of study. The definition for temperament appears to be more appropriate for use when reviewing the genetic parameters of behavioural differences in individuals and as such, during this thesis, the term temperament is used for consistency. However, when authors have used specific terms, these shall be referred to (e.g. if the study specifically states 'personality', this will be used). During the present study, the difference between temperament and personality in a guide dog population will be explored further.

1.2.1 Behaviour, temperament and personality in dogs

Animal behavioural and temperament research has centred around dogs due to their close relationships with humans, large numbers in the population and noticeable differences between breeds. There are numerous studies that report measuring behaviour (Hsu and Serpell, 2003; Haverbeke *et al.*, 2009; Rayment *et al.*, 2016), temperament (Goddard and Beilharz, 1984; Murphy, 1998; Hsu and Serpell, 2003; De Palma *et al.*, 2005; Duffy and Serpell, 2012) and personality (Svartberg and Forkman, 2002; Harvey, 2014; Harvey *et al.*, 2016) in dogs. Gosling

and John (1999) reviewed the human 'Big Five' of personality traits published by McCrae and John in 1992, reporting that extraversion, neuroticism and agreeableness demonstrated the strongest cross-species generality. However, it could be argued that constructs of personality cannot be directly compared between species; would the neuroticism of a dog be comparable to a chimpanzee for example? Following on from this work, Gosling *et al.* (2003) claimed that the five human personality traits could be compared in dogs. Four dimensions were reported to exist in dogs, 1) energy (analogous to extraversion), 2) affection (analogous to agreeableness), 3) emotional reactivity (analogous to neuroticism) and 4) intelligence (analogous to openness/intellect). Both the dog and human personality was scored by the dog's owner, so it is therefore not surprising that there are correlations reported between dogs and humans using this method.

Other researchers have delved deeper into dog personality traits and stepped away from the links to human personality. Svartberg and Forkman (2002) reported five different personality traits in dogs: playfulness; curiosity/fearlessness; chase-proneness; sociability; aggressiveness. All traits apart from aggressiveness were related to each other for eight breed groups (1175 dogs from 47 breeds). The study reviewed data collected in standardised tests (the 'Dog Mentality Assessment'), with the method noted to be less subjective than previous work due to using direct behavioural observations. The same measures and an additional shyness-boldness broader dimension were used by Svartberg *et al.* (2005) to investigate the consistency of traits. Playfulness, sociability, chase-proneness and boldness were consistent over the three repeated tests, with curiosity/fearlessness increasing and aggressiveness decreasing over the measures, suggesting not all traits were consistent over time. Following on from this work, one of the most well-known reviews of dog temperament and personality was published by Jones and Gosling (2005). The authors reviewed temperament literature and reported seven broad dimensions: reactivity, fearfulness, activity, sociability, submissiveness,

responsiveness to training and aggression. Studies have since reported numerous different dog temperament or personality traits. Ley *et al.* (2008) reported five underlying personality factors (factor 1 - degree to which the dog is energetic and extraverted, factor 2 - how self-assured and internally motivated a dog appears, factor 3 - how responsive the dog is to training, factor 4 - degree to which the dog is considered to be friendly and sociable, factor 5 - reflecting nervous sensitivity) based on adjectives used by owners to describe their dog. Data were collected from just over 1000 dog owners using similar methods to personality studies in humans. The terminology used to categorise the dogs' behaviour was subjective and open to interpretation, using terms such as 'intelligent', 'biddable', and 'dominant'. It could be argued that it is not possible to categorise dogs using human centred subjective terminology, and that the personality factors reported are not robust.

Following a review of dog temperament (or personality) traits within literature, the most frequently reported traits were aggression, body sensitivity, chase, distraction, energy/activity, excitability, fear/anxiety, playfulness, sociability and trainability (Table 1.1). The reference to these traits may be due to the specific focus on known and well reported traits, and the need to monitor behaviours for a specific purpose, such as those important to rehoming shelters or working dogs. The reference to these traits does not necessarily encompass all possible traits within dogs and there are likely to be more and higher categories of underlying traits. It is also unknown whether researchers are reporting the same traits. For example, whether all studies claiming to report 'fear', define and/or measure this in the same way. The data are commonly reported by different handlers (with varying levels of experience), using a variety of methods (e.g. questionnaires or practical assessments) for dogs of different breeds, ages and sex. When searching for 'fear in dogs', a range of papers are available. Fearful dogs in veterinary practices were noted to display lowered/tucked tails between the legs, staring ahead, crouched body posture, trembling or avoidance (Döring *et al.*, 2009). Stellato *et al.*

(2017) used a behavioural assessment, with fearful dogs showing avoidance behaviour, reduced body posture, displacement, vocalisations and subtle behaviours (body shaking, hiding, yawning), whereas Sheppard and Mills (2003) stated 14 behaviours associated with fear, including elimination, destructive behaviours and aggressive behaviours. Studies utilise questionnaires to measure fear (Duffy and Serpell, 2012; Blackwell *et al.*, 2013; Flint *et al.*, 2018a; Dollion *et al.*, 2019), but definitions are not always provided, e.g. Duffy and Serpell (2012) asks whether the dog “acts anxious or fearful”, and Dollion *et al.* (2019) asks if the dog has been “been wary or scared”, terms which can be open to interpretation. Therefore, it could be that these traits are more groupings of correlated behaviours for a specific population, rather than defined traits within the dog population.

Table 1.1 Temperament traits for dogs reported within the literature, indicating trait name and references.

Trait name	References
Aggression	Humphrey and Warner (1934), Bradshaw and Goodwin (1999), Serpell and Hsu (2001), Svartberg and Forkman (2002), Jones and Gosling (2005), Svartberg (2006), Duffy and Serpell (2008), Jones (2009), Takeuchi <i>et al.</i> (2009a), Mirkó <i>et al.</i> (2012), Wright <i>et al.</i> (2012), Curb <i>et al.</i> (2013), Arata <i>et al.</i> (2014), Foyer <i>et al.</i> (2014), Bálint <i>et al.</i> (2017), van der Borg <i>et al.</i> (2017), Brady <i>et al.</i> (2018), Clay <i>et al.</i> (2020), Wallis <i>et al.</i> (2020)
Body sensitivity	Humphrey and Warner (1934), Bartlett (1976), Goddard and Beilharz (1982), Murphy (1998), Harvey <i>et al.</i> (2016), Braem <i>et al.</i> (2017)
Chase	Serpell and Hsu (2001), Svartberg and Forkman (2002), Arata <i>et al.</i> (2014), Harvey <i>et al.</i> (2016), Plant (2016), Emily <i>et al.</i> (2024), Cooper <i>et al.</i> (2025)
Distraction	Goddard and Beilharz (1982), Takeuchi <i>et al.</i> (2009a), Arata <i>et al.</i> (2010), Plant (2016), Harvey <i>et al.</i> (2016), Harvey <i>et al.</i> (2017), Hare <i>et al.</i> (2024)
Energy/Activity	Humphrey and Warner (1934), Bartlett (1976), Serpell and Hsu (2001), Jones and Gosling (2005), Svartberg (2006), Russenberger and Clothier (2008), Takeuchi <i>et al.</i> (2009b), Mirkó <i>et al.</i> (2012), Harvey <i>et al.</i> (2017)
Excitability	Goddard and Beilharz (1982), Hsu and Serpell (2003), Bennett and Rohlf (2007), Svartberg (2006), Jones (2009),

	Hoffman <i>et al.</i> (2013), Bray <i>et al.</i> (2015), Harvey <i>et al.</i> (2016), Fallahi <i>et al.</i> (2024)
Fear/Anxiety	Goddard and Beilharz (1982), Serpell and Hsu (2001), (Diederich and Giffroy, 2006), Svartberg (2006), Jones (2009), Dreschel (2010), De Meester <i>et al.</i> (2011), Arata <i>et al.</i> (2014), Temesi <i>et al.</i> (2014), Harvey <i>et al.</i> (2016), Rooney <i>et al.</i> (2016), Bellamy <i>et al.</i> (2018), Barnard <i>et al.</i> (2019), Clay <i>et al.</i> (2020), Wallis <i>et al.</i> (2020), Bai <i>et al.</i> (2023)
Playfulness	Svartberg and Forkman (2002), Svartberg (2006), Jones (2009), De Palma <i>et al.</i> (2005), Barnard <i>et al.</i> (2019), Lampe <i>et al.</i> (2019), Svartberg (2021)
Sociability	Svartberg and Forkman (2002), De Palma <i>et al.</i> (2005), Jones and Gosling (2005), Svartberg (2006), Russenberger and Clothier (2008), Mirkó <i>et al.</i> (2012), Riemer <i>et al.</i> (2013), Arata <i>et al.</i> (2014), Persson <i>et al.</i> (2015), Shin and Shin (2017), Brady <i>et al.</i> (2018), Barnard <i>et al.</i> (2019), Svartberg (2021)
Trainability	Pfaffenberger and Scott (1959), Bradshaw and Goodwin (1999), Serpell and Hsu (2005), Ley <i>et al.</i> (2008), Marshall-Pescini <i>et al.</i> (2008), Jones (2009), Kubinyi <i>et al.</i> (2009), Helton (2010), Turcsán <i>et al.</i> (2011), Mirkó <i>et al.</i> (2012), Harvey <i>et al.</i> (2017), Nose and Kakinuma (2019), Deshpande <i>et al.</i> (2024)

Breed contributes to the expression of behaviour, and thus temperament traits displayed by dogs (Svartberg and Forkman, 2002; Mehrkam and Wynne, 2014). Breed differences have been reported in many studies, including Ley *et al.* (2009), Turcsán *et al.* (2011), Mirkó *et al.* (2012), Schneider *et al.* (2013) and Sundman *et al.* (2016). In comparison, some studies have reported few differences in behaviour between breeds. Svartberg and Forkman (2002) found no or little differences for curiosity/fearlessness, chase-proneness and aggressiveness for individual breed groups, but some differences for playfulness and sociability. Similarly, Svartberg *et al.* (2005) found no difference in traits by breed groups when testing consistency of personality traits in three breed groups (herding dogs, guarding dogs and gun dogs), however the sample sizes for all three groups were 10 or below. Studies between and within breeds have focussed on set populations with specific traits of interest and therefore may explain some of the differences reported. A wide cross-

sectional review of dog behaviour is needed to understand true differences between breeds. Reviewing the genetic influences on behaviour between breeds will also widen this understanding.

Dog temperament can influence whether dogs are able to perform roles such as working or assistance dogs. Subsequently, many temperament studies on dogs have been carried out on working and assistance dogs to aid selection for breeding or training (Jones and Gosling, 2005; Brady *et al.*, 2018) and to determine the likelihood of success within the working role (Serpell and Hsu, 2001; Batt *et al.*, 2008; Harvey *et al.*, 2017).

Working dog organisations typically have data available spanning many generations, with standardised socialisation, rearing and training practices. Therefore, the variation in environment, upbringing (Vaterlaws-Whiteside and Hartmann, 2017) and training methods used are more consistent than within pet dog populations, enabling examination of behaviour with less influence from confounding variables. However, this method of upbringing may limit exposure to certain stimuli, and so the response of individual dogs to stimuli they have not experienced would not be known or could be predicted. There are specific behaviours that are important to measure and select for in working dog populations that may not be considered so important in pet dog populations (Hunt *et al.*, 2020). Brady *et al.* (2018) systematically reviewed the behavioural characteristics important to working dogs, reporting six temperament traits: approach-withdrawal; human-directed aggression; human-directed social behaviour; object-directed play; sensitivity to aversities, and willingness to work. These traits are different to the most reported temperament traits within the literature (Table 1.1), further highlighting that traits of importance vary depending on the role of the dog. Even within working dogs there are variations in selection for certain behaviours depending on the specific working tasks required. For example, studies using scent detection dogs and military dogs frequently report gun-shyness and response to sudden and random items such as pop up ghosts (Wilsson and Sundgren, 1997; Svartberg and Forkman, 2002), which are appropriate for the working environment of the

dogs. Other examples include guide dogs, where traits such as distractibility, attachment and body sensitivity are important as dogs need to be able to focus on guiding whilst withstanding a harness and handling by clients with a vision impairment (Hunt *et al.*, 2020).

Several studies have investigated dog behaviour in guide dog populations. One of the first to report behavioural traits in guide dogs identified 10 traits in a USA population: body sensitivity; ear sensitivity; protective instinct; fighting instinct; confidence; self-right; energy; willingness; intelligence and nose acuity (Bartlett, 1976). Following this work, there was little published information until Goddard and Beilharz (1983), where five temperament traits were reported in a population of 1031 Australian guide dogs developed from behavioural withdrawal reasons: distraction, general performance, sensitivity, fearfulness and fearfulness accompanied with high activity. Since 1983, numerous papers have reported traits in guide dogs (Serpell and Hsu, 2001; Batt *et al.*, 2008; Schiefelbein, 2012; Harvey *et al.*, 2016; Harvey *et al.*, 2017), and the heritability of those traits (Evans *et al.*, 2015). The traits commonly reported in guide dog populations include distraction (Goddard and Beilharz, 1983; Harvey *et al.*, 2016), fear (Goddard and Beilharz, 1983; Schiefelbein, 2012), excitability (Goddard and Beilharz, 1983; Schiefelbein, 2012), energy (Bartlett, 1976; Harvey *et al.*, 2016) and aggression (Takeuchi *et al.*, 2009a; Schiefelbein, 2012). Due to selective breeding, little evidence of aggression in guide dog populations has been found (Goddard and Beilharz, 1983; Harvey, 2014; Hunt *et al.*, 2020).

1.2.2 Methods to monitor behaviour and temperament in dogs

Within the scientific literature there are multiple published and validated methods for measuring behaviour and temperament within a range of species, with studies often conducted in social species such as dogs. Temperament is commonly measured by either behavioural coding or rating (Mehta and Gosling, 2008). Behavioural codings are commonly

reported as behavioural observations (Wilsson and Sundgren, 1997; Sinn *et al.*, 2010; Asher *et al.*, 2013) and ratings tend to be owner or trainer reported questionnaires (Hsu and Serpell, 2003; Mirkó *et al.*, 2012; Harvey *et al.*, 2016), or a mixture of both (Schneider *et al.*, 2013; Bellamy *et al.*, 2018). McGarrity *et al.* (2016) found no difference in the predictive validity of behavioural ratings and codings when measures were compared in a sample of 52 explosive detection dogs with equal numbers of males and females from three breeds. Given the small number of individuals in the study, extrapolation of results to all working dogs in general may not be appropriate. The authors do not refer to a power analysis to confirm an adequate sample size and were restricted by the number of dogs available with video footage for analysis.

The age of dogs and time points at which behavioural information is collected are important considerations. Harvey *et al.* (2016) found the most appropriate time points to collect behavioural information from guide dog puppies to be five, eight and 12 months of age, with eight and 12 month data shown to be more stable and predictive of later success. In other studies, the age of behaviour data collection varies from as early as six weeks (Asher *et al.*, 2013), during the juvenile phase (Duffy and Serpell, 2012; Harvey *et al.*, 2016) and into adulthood (Mirkó *et al.*, 2012). Temperament consistency has been a focus of research in dogs (Choi *et al.*, 2017) and evidence suggests that dogs exhibit more consistency as they age (Goddard and Beilharz, 1986). Svartberg *et al.* (2005) reported that temperament consistency in adult dogs was observed over an interval of one to two years in a cohort of 40 dogs. Chopik and Weaver (2019) found that fearfulness and aggression were similar between younger and older dogs, however activity/excitability was greater for younger dogs, suggesting that not all factors remain consistent over the dog's lifetime. In a study investigating the factors associated with success in guide dog training, behavioural traits measured at six months were found to be relatively unstable, whereas those measured at 14 months were better predictors of success (Batt *et al.*, 2008).

1.2.2.1 Questionnaires

A common method for collecting behavioural data from a large and geographically wide-spread population, such as for pet dogs, is by using questionnaires. An owner-completed questionnaire has advantages as owners typically know more about their dog's usual behaviour than an individual not living with the dog (Serpell and Hsu, 2001). However, owner-reported questionnaires can be open to high levels of bias; individuals can interpret and view dogs differently dependent on human-animal bond, experience and skills (Bellamy *et al.*, 2018). Owners may also be more likely to report behavioural problems in their dogs when survey responses are kept anonymous (Segurson *et al.*, 2005). To monitor dogs within populations or working dog programmes it is not possible for responses to be kept anonymous as individual dog progression must be monitored. There is therefore the potential for the withholding of information or under-reporting of behaviours. It is possible that this may occur less in working dogs as volunteers commonly raise the dogs and are aware of the importance of reporting behaviour accurately. It also may not always be possible for individuals to detect some of the less obvious behavioural signs, and as a result there is a possibility of false negatives within any behavioural data (Overall *et al.*, 2016). Volunteers raising assistance or working dogs are commonly provided with training on dog behaviour, so it may be that this population are able to recognise more subtle behaviours than pet dog owners, but this theory is not evidence-based.

As well as owners, trained professionals can also individually influence scores reported in behavioural questionnaires. As part of the Epidemiology of Guide Dogs study undertaken by the University of Nottingham, training staff accounted for a considerable proportion of variation in behavioural reports (Asher *et al.*, 2016). There have however been reports of high correlations between trained and untrained experts' reports of dog behaviour. Fratkin *et al.* (2015) reported that minimally trained and non-expert individuals were able to assess the behaviour of military working

dogs and explosive-detection dogs to an acceptable standard when compared to trained experts. This is suggestive that similar biases may be present in any data collected by individuals that know the dog or that biases differ between organisations.

One of the most widely used behavioural questionnaires is the Canine Behavioral Assessment and Research Questionnaire (C-BARQ; Hsu and Serpell, 2003). The C-BARQ is widely used amongst different dog populations, including shelter, working and pet dogs (Serpell and Hsu, 2005; van den Berg *et al.*, 2006; Duffy and Serpell, 2012; Jakuba *et al.*, 2013; Foyer *et al.*, 2014; González-Ramírez *et al.*, 2017) and measures 14 traits (stranger-directed aggression; owner-directed aggression; dog-directed aggression; dog rivalry; stranger-directed fear; non-social fear; dog-directed fear; separation-related behaviour; attachment and attention-seeking; trainability; chasing; excitability; touch sensitivity; energy level). The suitability of the C-BARQ for use with some working dog populations has been questioned due to the importance of measuring specific behaviours in working dogs that are omitted from the C-BARQ, such as distractibility (Foyer *et al.*, 2014; Hunt *et al.*, 2020). Yet there are direct comparisons between the two questionnaires, with 22 of the 39 r-PWQ questions originating from C-BARQ questions and seven of the eight corresponding traits were shown to be significantly correlated (Hunt *et al.*, 2020).

Due to the potential for general associations between similar animals and possible errors, data may be more accurate when collected from several individual observers (Goddard and Beilharz, 1983; Bellamy *et al.*, 2018). The bias from behaviour questionnaires is hard to limit, however if clear terminology is used in questions, then the difference in interpretation of questions can be reduced. The greatest agreement between trainers was reported for traits with well-defined responses (Goddard and Beilharz, 1983). Another important factor to consider within behavioural questionnaires is the scale. Likert scales are commonly used to represent

levels of a trait (Meagher, 2009), however variation can be caused by the participants' interpretation of the terms used and there is less sensitivity in the scales (Cummins and Gullone, 2000). Visual analogue scales can detect smaller changes and have been reported to be reliable for assessing the level of pain in dogs (Hielm-Björkman *et al.*, 2011) as well as for measuring puppy behaviour (Harvey *et al.*, 2017). Previous research found that guide dog trainers often found it difficult to categorise dog behaviour using a Likert scale and that it may not always be easy to place a dog within a category (Harvey, 2014). Goddard and Beilharz (1983) reported differences in behaviour reports between guide dog trainers. The study used a 3-point Likert scale, so difference in opinion had a considerably greater impact on the scores. The use of a wider visual analogue scale, such as that used in a puppy behaviour questionnaire for guide dogs (the PWQ; Harvey *et al.*, 2016) potentially reduces the impact of this due to the large range of scores. However, variation between trainers will always be present, and is something that is difficult to control in a large geographically widespread programme. Whilst questionnaire data have some potential for bias, collecting data in this way enables large-scale studies to be undertaken with many participants and the ability to review genetic variance in populations (Ilska *et al.*, 2017).

1.2.2.2 Behavioural assessments

Behavioural assessments are undertaken in a variety of different dog populations, including with assistance dogs due to the number of individuals available for assessing. Assessments have been developed for purposes such as assessing shelter dogs (Valsecchi *et al.*, 2011), determining suitability of assistance dogs (Tomkins *et al.*, 2011) or working dogs (Lazarowski *et al.*, 2021), and to understand traits present in pet dogs (Svartberg, 2021). Between studies and assessments there are differences in sample sizes, age at assessment, tasks measured, and data reported.

Assessments have been undertaken in young assistance dog puppies, between six and eight weeks of age. Asher *et al.* (2013) developed a behavioural test in a cohort of 587 guide dog puppies, and the assessment was later shown to demonstrate heritabilities in over 2500 puppies (Evans *et al.*, 2015). The same assessment was then undertaken in a cohort of 249 guide dog puppies in the USA where no subtests were associated with guide dog success (Feng *et al.*, 2024). In a different cohort of assistance dogs, Bray *et al.* (2020) assessed 168 puppies for cognitive characteristics, with some of the human-related assessments showing heritabilities (Bray *et al.*, 2021a). Other assessments have been undertaken in older dogs. Arvelius *et al.* (2014) analysed data for 873 German Shepherd Dogs bred by the Swedish Armed Forces between 11 and 23 months of age, with seven aggregate behaviour scores shown to have a heritable aspect. Svartberg (2021) reported results for a larger sample, over 12,000 dogs older than 12 months of age, using the Behaviour and Personality Assessment in Dogs (BPH). The BPH consists of eight elements and measures aspects of 'boldness', which is a higher trait that includes other behavioural factors such as sociability/playfulness and non-social fearfulness. Behavioural assessments have been trialled by guide dog organisations, with some smaller organisations able to undertake standardised assessments for all dogs in the programme.

Taylor and Mills (2006) also encouraged researchers to adopt a more scientific approach when developing behavioural assessments and highlighted the importance of standardisation, reliability, validity and practicality in a systematic review of behavioural assessments in working dogs. Brady *et al.* (2018) stated there was a large range of assessments used to measure dog behaviour, and noted how different terminologies were used for similar traits. It is therefore difficult to make comparisons between studies, and many studies failed to report the reliability and validity of the assessments. It would appear there are still issues present when using behavioural assessments.

1.2.3 Guide dog behaviour monitoring

Behaviour is reported to be the most frequent reason for withdrawal of dogs amongst guide dog schools (Goddard and Beilharz, 1982; Serpell and Hsu, 2001; Arata *et al.*, 2010; Blythe, 2020). Identification of behaviours associated with withdrawal is of great benefit to guide dog organisations in terms of minimising costs, maintaining production efficiency and high welfare standards. Studies have reviewed the reasons for removal from guide dog programmes (Goddard and Beilharz, 1982; Goddard and Beilharz, 1983; Serpell and Hsu, 2001; Evans, 2014; Harvey, 2014; Asher *et al.*, 2016; Caron-Lormier *et al.*, 2016; Blythe, 2020) with an aim to increase success rates. Goddard and Beilharz (1982) reported the most common traits for unsuccessful guide dogs in Australia to be fearfulness, distraction and aggression, with an average success rate of 47%. Ennik *et al.* (2006) described the graduation rate at The Seeing Eye (USA) to be 56% with 39% of dogs needing to undertake the training cycle more than once. Guide Dogs UK (GDUK) has stated qualification rates to be as high as 70% (Evans, 2014). However more recent unpublished data monitoring progression of puppies born each year found qualification rates to range from 50-60% for dogs born between 2013 and 2016, with approximately 75% of dogs withdrawn due for behaviour reasons (Moxon, 2018). When reviewing the main reasons for withdrawal, stress resilience, distraction and suspicion/aggression are the most frequently reported reasons (Serpell and Hsu, 2001; Evans, 2014; Moxon, 2018; Blythe, 2020). Similarly, Caron-Lormier *et al.* (2016) reported the main reasons for early retirement of guide dogs (retirement indicates the end of a working partnership, rather than withdrawal where a dog is removed from working as they are deemed unsuitable for the role based on health or behaviour) to be environmental anxiety (mainly stress resilience), willingness/confidence and fear/aggression (mainly suspicion).

Many guide dog schools use questionnaires or assessments to measure the behaviour of dogs during puppyhood and before training commences in order to make decisions on dogs (Batt *et al.*, 2009; Asher *et al.*, 2013;

Harvey *et al.*, 2017). Harvey (2014) developed the Puppy Walker Questionnaire specifically for use in guide dog populations due to the importance of certain behaviours. A refined version of the questionnaire is now used by GDUK (the refined Puppy Walker Questionnaire; r-PWQ; Hunt *et al.*, 2020). The questionnaire consists of 39 questions that form nine components considered important to guide dog populations (separation; excitability; general anxiety; animal chase; trainability; body sensitivity; attachment; energy; distractibility). Component scores have been linked to outcome as a guide dog in a population of 214 puppies at eight months and 226 at 12 months of age (Harvey, 2014), then again in a population of 359 puppies at eight months of age (Hunt *et al.*, 2020). During development, the questionnaire demonstrated some sire and litter effects (Harvey, 2014), suggesting there may be a heritable component.

GDUK have previously trialled behavioural assessments at different time points. Data for 587 six- to eight-week-old puppies born between 2006 and 2007 showed associations to guide dog success in five stimuli using Puppy Profiling Assessments (Asher *et al.*, 2013) with heritabilities for the assessment ranging from 0.02 to 0.25 (Evans *et al.*, 2015). Harvey *et al.* (2016) then undertook a juvenile puppy test for 93 puppies from 2011 to 2012. Puppies were assessed at five and eight months of age with some assessment items associated with guide dog success. Only one item was significantly associated with training outcome at both time points: time oriented towards food. Feasibility for using the assessment at GDUK was explored, but due to the geographical spread of puppies in puppy raising, and the need to accurately assess and standardise the assessments, the assessment was not implemented for business as usual (BAU). Blythe (2020) developed CADA (Controlled Adult Dog Assessment) for 111 dogs in training during 2013, when dogs were around 14 months of age. There were no significant associations found between the CADA scores and guide dog success. The assessment was not implemented by GDUK as BAU.

It is important to note that the measure of guide dog success can vary between organisations and assessments and often subjective selection/assessments by trainers are used (Murphy, 1998). There are also many other factors that can influence the outcome of guide dogs, including but not limited to; breed, gender, neuter status, health conditions, dog and trainer availability, dog flow, demand for guide dogs and client matching criteria (Murphy, 1995). However, there are issues controlling for such factors as they cannot be consistently and accurately determined. Due to the large number of dogs unable to succeed as guide dogs for behavioural reasons, it would be advantageous if breeding programmes were able to utilise selection for behaviour using measures such as estimated breeding values (EBVs). GDUK has previously trialled using subjective behavioural measures to create EBVs for behaviour, which were unsuccessful. Since then, the r-PWQ was introduced as BAU to collect information on guide dog puppies (Hunt *et al.*, 2020).

1.3 Breeding selection

Breeding programmes aim to improve and expand groups of animals for the next generation dependent on certain criteria. This is commonly based on the purpose of the animals being bred, whether for farming purposes, as companion animals, as working animals or to protect/increase a species. Selection for more profitable or successful animals is a main objective of most breeding programmes (Andersson, 2001; Bohmanova *et al.*, 2010). Decisions on which, and how many, animals to breed can also influence animal health and welfare (Czerwinski *et al.*, 2016). Breeders therefore have a responsibility to ensure selection of the most appropriate animals. Animals can be selected on phenotypic and genetic information and dependent on the environment and resources available. Some breeding programmes are now able to select animals with desirable behaviours. Studies in minks have described selecting for low stereotypic behaviour (Svendsen *et al.*, 2007) and against animals that present fur chewing (Malmkvist and Hansen, 2001). Within farming, breeding for

behaviour has been introduced to improve production and animal welfare, examples include selecting cattle for ease of handling (D'Eath *et al.*, 2010).

Commercial dog breeding has focussed heavily on breeding for appearance and human demand, often compromising the welfare and health of dogs through excess breeding, puppy farming and breeding dogs with known health issues. There have been many debates over the ethical nature of pedigree dog breeding (Rooney and Sargan, 2009; Bovenkerk and Nijland, 2017). A BBC-documentary in 2008 titled 'Pedigree Dogs Exposed' raised public awareness of the issues caused by some dog breeding practices (BBC, 2008). Following this, Bateson (2010) produced an independent enquiry into dog breeding and recommended revising the Kennel Club breed standards to reduce selection of extreme morphological qualities. Two main problems have been identified with pedigree dogs, first the selective breeding for extreme looks (that are often associated with a number of health issues) and second the narrow gene pool within some breeds that leads to loss of genetic diversity due to intense selective breeding (Bovenkerk and Nijland, 2017; Keijser *et al.*, 2018b).

1.3.1 Phenotypic selection

The mating of two individuals will result in a combination of both parents' genes and will influence the phenotype presented by their offspring (Falconer and Mackay, 1989). Phenotype is the observable characteristics of an individual, which are influenced by both genotype and the environment (Mackenzie *et al.*, 1986). Behavioural phenotypes can be measured in a variety of ways (Overall *et al.*, 2014), with observations and questionnaires most commonly used for dogs (Mehta and Gosling, 2008). The methods used to measure behaviour and temperament in dogs have been discussed above (see 1.2.2).

Historically, breeding programmes have selected animals based on the diagnosis of health conditions, phenotypic displays of behaviour and appearance. The phenotypic selection of animals by humans has diversified the appearance and domestication of many species (Andersson, 2001). However domesticating animals can cause populations to change genetically to enable them to adapt to new environments (Price, 1998). Over the years, the understanding of genetic influences on phenotype and the availability of genetic screening has influenced breeding decisions. Thus, many breeding programmes do not now rely on phenotype information alone when planning a mating. To accurately select for temperament traits, an understanding of the genetics that influence the behavioural traits is required (Mackenzie *et al.*, 1986).

1.3.2 Genomic selection

Genetics is an ever-growing field of science, covering a range of species, delving deeper into complex and common health conditions and linking to behavioural presentation. The development of DNA-based markers has revolutionised animal and plant genetic study capabilities (Dodgson *et al.*, 1997; Vignal *et al.*, 2002). Single Nucleotide Polymorphisms (SNPs) are DNA-based markers commonly referred to within genetic studies (Vignal *et al.*, 2002) and account for a large proportion of human genetic variation, allowing for the mapping of complex traits (Marth *et al.*, 1999). A SNP is a single code change within a single base pair of DNA (Grandin and Deesing, 2013). SNPs are a major factor in genetic variation and are involved in disease inheritance. Research into the genetics of canine diseases has helped support the understanding of human diseases due to comparable genetic structure (Ostrander and Williams, 2019). It is thought that up to 350 diseases are found in both dogs and humans (Hayward *et al.*, 2016). Improvements in sampling method quality and affordability has advanced the field and capabilities of genetics (Dunn and Ryan, 2015).

Focussing on the whole genome of an animal (often termed genomics), rather than a few individual genes, gives researchers the capability to study DNA more comprehensively (Dodgson *et al.*, 1997). The affordability of genotyping and improved technology allows breeding programmes to utilise genomic selection more widely (Ma *et al.*, 2018). Genetic sequencing and tests have been able to support breeding decisions in a range of species including for farm animals such as cattle (Hayes *et al.*, 2009) and pigs (Bovo *et al.*, 2020), where selection has focussed traits of economic importance such as animal growth, meat quality and milk production (Andersson, 2001). It is important to remember that not all genome changes will result in changes in phenotype (Dunn and Ryan, 2015), particularly when complex diseases and behavioural traits are studied as often a number of genes, along with environmental influences, are involved in the presentation of complex traits.

A better understanding of the genetic basis for behavioural phenotypes can inform breeding programmes (Goddard and Beilharz, 1982; Joshi, 2005; Grandin and Deesing, 2013). Breeding for behaviour can improve animal welfare by selecting animals with more desirable behaviours. Should the behaviours have a genetic influence, and be selected appropriately, more desirable behaviours could be seen in progeny. This could reduce the number of companion animals relinquished or euthanised due to behavioural reasons (Jensen *et al.*, 2008; D'Eath *et al.*, 2010). However, breeding for behaviour can often be more challenging than breeding for physical traits or health as behavioural traits are commonly affected by multiple genes (Plomin, 1990; Evans *et al.*, 2015). There have, however, been associations found between specific genes and the display of complex behaviours. An example is the association between the DRD4 receptor and novelty-seeking behaviour in birds, suggesting the gene influences variation seen within avian and mammalian lineages (Fidler *et al.*, 2007).

Dogs have been a focus in the genomic field due to their close relationship with humans and vast morphological differences found between breeds. Scientists have sequenced over 2.5 million SNPs in the domestic dog (Lindblad-Toh *et al.*, 2005). The availability of genetic sequencing technology has resulted in an increase in dog genetic studies, including larger scale projects such as the Working Dog project (Working Dog Project, n.d.). Canine genetic analyses have demonstrated associations for neoplasia, cardiovascular disease, immunologic disorders and metabolic disease (Ostrander and Williams, 2019). Using 4224 samples, Hayward *et al.* (2016) identified seven significant quantitative trait loci associated with lymphoma and idiopathic epilepsy. Lavrijsen *et al.* (2014) reported patellar luxation in a study of 93 Flat Coated Retrievers (45 affected and 48 controls) to be significantly associated with a region on Chromosome CFA07, with a suggestion that a genetic test for the breed could be developed. DNA testing has already been developed for some more prevalent breed-specific diseases in dogs, including for German Shepherd Dogs, Golden Retrievers and Labrador Retrievers (Czerwinski *et al.*, 2016), breeds that are commonly used as guide dogs. DNA tests are available for diseases such as degenerative myelopathy, progressive retinal atrophy, dwarfism, ichthyosis and exercise-induced collapse (Czerwinski *et al.*, 2016), which are used within the GDUK breeding programme (W. Adams 2020, *personal communication*, 12 August 2020). A recently published study reported differences in genomic prediction models and the possibility for genomic breeding values (GBVs) in a population of USA guide dogs (Thorsrud *et al.*, 2025). Seeing Eye evaluated three binary health traits (anodontia, distichiasis, oral papillomatosis) and one behavioural trait (distraction), concluding phenotypes could be selected for using GBVs.

In order for genetic studies to be useful and reliable, the phenotypic information analysed within the study needs to be robust and reliable, and there needs to be a large enough data set to highlight differences (Liinamo *et al.*, 2007; Iliska *et al.*, 2017; Friedrich *et al.*, 2019). For a reliable

quantitative genetic analysis, it is recommended that at least 500-1000 dogs with known phenotypes be included, and that all descendants of the sire and dam be studied if possible (Liinamo *et al.*, 2007). It is also essential that the parentage is recorded accurately; inaccurate parentage information can result in the incorrect assignment of 'hereditary clear' statuses (Lewis, 2019).

1.3.3 Heritability

Genetic correlation studies commonly report heritability (often denoted as h^2), which indicates the probability of a specific trait or health condition changing through breeding selection within a population (Falconer and Mackay, 1989). The heritability of a trait (whether health or behavioural) is described using a scale of 0 to 1, representing the proportion of variance in the trait attributed to genetics, rather than environmental factors (D'Eath *et al.*, 2010). The higher the number, i.e. closer to 1, the higher the heritability of the trait (Lewis, 2016). Heritability quantifies the relationships between phenotypic and genotype information (Ginja *et al.*, 2008) and is important to consider when making breeding decisions (Lush, 1949). However, some caution is required as these values are only estimates based on the sample and data available (Falconer and Mackay, 1989).

In farm animals, where a large proportion of work on genetic and breeding selection began to improve productivity and to select against undesirable health/behaviour (Rodenburg and Turner, 2012), heritabilities are reported to range depending on the factors measured and the populations studied (Falconer and Mackay, 1989). The heritability of behaviour and health in dogs is well reported in the literature, with differences found between breeds and populations. For example, the heritability of hip laxity scores was 0.83 in a population of 215 Estrela Mountain Dogs (Ginja *et al.*, 2008), compared to a heritability of 0.26 for hip dysplasia in a population of 575 German Shepherd Dogs in the USA (Mackenzie *et al.*, 1985). Heritability estimates for behavioural traits in dogs are generally reported to range

from 0.0 (Wiener *et al.*, 2015) to 0.8 (Liinamo *et al.*, 2007), however the majority of heritability estimates are reported to be under 0.4 (Ilska *et al.*, 2017). It is suggested this may potentially be due to the effects of learning and training on behavioural displays (Mackenzie *et al.*, 1986).

Breed differences in behavioural patterns appear to be consistent across studies (Serpell and Hsu, 2005; Mehrkam and Wynne, 2014). It is therefore not surprising that differences in behavioural heritability estimates are found between breeds. Wilsson and Sundgren (1997) reported German Shepherd Dogs ($n=131$) had high heritability for affability (defined as the dog's willingness to contact people, scored on a 1-12 scale; $h^2=0.37$) in comparison to Labrador Retrievers ($n=797$, $h^2=0.15$), despite both being bred and raised for similar purposes (as either service or pet dogs). Hayward *et al.* (2016) also reported that if single breeds are studied, a large amount of genetic variation can be missed that may influence phenotype prevalence. MacLean *et al.* (2019) noted that measures between breeds yield higher heritability estimates, averaging five times higher than within breed estimates due to wider genetic variation seen between breeds. Studies therefore more commonly report between rather than within breed differences.

Behavioural questionnaires have been used for heritability analysis. A recent study including over 14,000 dogs from over 100 breeds found within and between breed differences for 14 C-BARQ traits, with heritabilities ranging from 0.27 to 0.73 (MacLean and Bray, 2019). The traits with the highest heritability were trainability (0.73), stranger-directed aggression (0.68), chasing (0.62) and attachment and attention-seeking (0.56). The authors did not report controlling for breed within analyses but stated that 25 individuals from each breed were randomly selected and thus breeds with less than 25 observations were not included. Differences between breeds have also been reported for behaviours measured using the C-BARQ. Trainability heritability varied from 0.15 in population of around 2000 Labradors (Ilska *et al.*, 2017) to 0.25 for a population of 3591

dogs from a wider range of 20 breeds (Asp *et al.*, 2015). The heritability of aggression-related behaviours ranged from 0.77 in a population of 325 Golden Retrievers (Liinamo *et al.*, 2007) to 0.0 in around 2000 Labradors (Ilska *et al.*, 2017), and others have reported similar low heritabilities of around 0 for C-BARQ traits in different populations (Schiefelbein, 2012; Friedrich *et al.*, 2019). The difference in heritability estimates for dogs measured using the same questionnaire supports the need for regular monitoring of heritability estimates and an awareness that heritability can differ greatly within and between breeds and populations.

1.3.3.1 Pedigree-based heritabilities

Breeding programmes aim to select for both health, behaviour, or both, often by utilising breeding values. The breeding value of an individual for a specific trait is measured by the mean value of that trait for its progeny (Falconer and Mackay, 1989). Estimated breeding values (EBVs) are calculated from the phenotype of an animal, its relatives and pedigree relationships (Soo and Worth, 2015). As a breeding value expresses traits passed on to offspring from parents, the expected breeding values of new progeny is half the value of the father's plus the mother's breeding value and is a combination of both phenotype and heritability of selected factors or traits (Falconer and Mackay, 1989). As such, EBVs are more accurate measures to understand genetic influences rather than using phenotype information alone (Lewis, 2016). It is possible to predict the likely phenotype of an individual as early as birth, based on the breeding value of the parents (Lewis *et al.*, 2010). As more data become available on an individual and their relatives (parents, offspring and siblings), EBV accuracy improves (Lewis *et al.*, 2010).

Breeding values must be calculated for individual populations as the results may vary between populations (Janutta *et al.*, 2006). A population is an aggregate of cases, animals, people etc., as defined by the statistician to represent a specific need (Swinscow, 1997). Consequently, the population must be reported for each study or analysis, so this is clear

to the reader. It is not always possible or appropriate to translate findings from one population to another depending on how comparable the populations are. For example, using different behavioural measures or training techniques in assistance dog populations despite using dogs bred for a similar purpose, as demonstrated by the differences in heritability between guide dog schools reported by Schiefelbein (2012). The genetic potential of an animal can be increased by combining predictive breeding values of important traits in a selection index (Hazel, 1943). Selecting breeding animals based on genetic merit has the ability to accelerate the improvement of phenotype displays and rate of genetic gain within a population (Evans *et al.*, 2015; Soo and Worth, 2015). Calculating EBVs is a considerably cheaper and more cost-effective method than DNA testing animals and thus can be utilised more widely.

EBVs have been used in the farming industry within a range of species such as cattle (Rensing *et al.*, 2009) and pigs (Mallard *et al.*, 1992). EBVs have also been used for several years to monitor hereditary diseases within dogs. The UK Kennel Club and other large dog breeding programmes actively utilise EBVs to reduce the incidence of certain health conditions such as canine hip dysplasia and elbow dysplasia (Lewis *et al.*, 2013). Both of these diseases are known to have genetic influences (Janutta *et al.*, 2006) and cause considerable welfare implications for dogs (Ginja *et al.*, 2008). Lewis *et al.* (2013) demonstrated that utilising EBVs to select for hip and elbow dysplasia in 15 UK dog breeds resulted in scores being between 1.16 and 1.34 times more accurate when compared to using raw scores. Therefore, by selecting dogs for breeding with low EBVs, it is hoped that the prevalence in the population can be controlled and eventually reduced, although it can take at least two or three generations to see measurable genetic improvements (Leighton, 2009).

1.3.3.2 Genomic-based heritabilities

Genome-wide studies can deliver more accurate estimates of genetic influences than pedigrees alone (Keijser *et al.*, 2018b). Therefore, the use

of EBVs has advanced and, in combination with DNA information, studies have reported the use of genomic estimated breeding values (GBVs). Genomic information such as quantitative trait locus (QTL) mapping are used to calculate the breeding value of an animal based on the genetic information, resulting in GBVs rather than EBVs where pedigree information is utilised (Moser *et al.*, 2009; Calus, 2010). Similar to EBVs, GBVs started to be used by the farming industry. Hayes *et al.* (2009) determined that GBVs could be used to reduce the greenhouse gas emissions from Angus cattle by five percent over a period of 10 years. Ma *et al.* (2018) noted the use of GBVs for factors such as milk yield and fat percentage in a range of French bulls from different countries, supporting the idea that GBVs could be calculated from related populations providing the populations were large and animals were genetically similar.

In dogs, genomic heritabilities are suggested to be closer to 0 when compared to pedigree-based heritability estimates (Ilska *et al.*, 2017), potentially due to the relatively small sample size for genotyped dogs compared to those with pedigree estimates. This is common in genotype studies due to the expense of the DNA extraction which limits sample sizes. However, the accuracy of GBVs will be significantly greater than EBVs because actual genetic impact is being considered rather than estimated genetic impact. GDUK has recently launched the Born to Guide project, which aims to fully genome sequence 3000 guide dog puppies (Guide Dogs, 2020b). It is hoped that in the future GBVs will be possible within this guide dog population.

1.3.4 Behavioural genetics

Behavioural genetics studies are commonly undertaken in dogs. This is due to a wide variation of behaviour between dog breeds, the ability of dogs to undertake a range of working roles and the relationships dogs have developed with humans (Overall *et al.*, 2014). The great variation in dog phenotype provides opportunities for genetic analysis between and

within breeds (Åkerberg *et al.*, 2012). However, the impacts of genetics, environment and experience all influence the behavioural variation seen in dogs (Mackenzie *et al.*, 1986; Åkerberg *et al.*, 2012; Friedrich *et al.*, 2019), and as a result it can be difficult to describe the true phenotypic behaviour (Overall *et al.*, 2014). If associations can be identified between behavioural phenotypes and genetic variants, dogs can be selected for breeding to improve behaviour (Bellamy *et al.*, 2018). However, breeding for behavioural traits will only be of use to breeding programmes if the traits are measured accurately and they demonstrate genetic variation (Ruefenacht *et al.*, 2002).

Genome-wide association studies (GWAS) and QTL mapping have been used to determine associations between individual behaviours in animals including dogs (Joshi, 2005; Persson *et al.*, 2016; Iliska *et al.*, 2017). For behaviour, it is common to report genetic variance and heritability estimates for traits rather than for individual behaviours (Iliska *et al.*, 2017). MacLean *et al.* (2019) studied a population of 5697 dogs from 101 breeds and identified 131 SNPs associated with breed differences in behavioural traits expressed. However, the study failed to control for the age of the dog at C-BARQ completion. Age can impact the phenotype displayed based on numerous factors including development stage, learnt behaviour, environment change and health conditions; therefore, the findings should be interpreted with caution. Recently, Friedrich *et al.* (2019) reviewed the heritable behaviour of 1041 German Shepherd Dogs using C-BARQ questionnaire responses in conjunction with GWAS and regional heritability mapping (RHM). Several genes associated with specific behavioural traits and correlations between pedigree and genome-based heritability estimates were reported. The study population included dogs aged over two years, reducing the influence of some confounding variables such as juvenile development. However, again, the authors do not report the age range of the dogs at C-BARQ data collection. Additionally, the dogs were from multiple countries, therefore differences in environment

and experience, along with cultural differences, that were not controlled for in the analysis could have affected the behavioural data used.

In recent years, studies have begun looking into whether specific genes are related to behavioural traits and phenotypes in dogs. Examples include the work by Persson *et al.* (2016), where two genomic regions (in the SEZ6L gene and ARVCF gene) were associated with human-directed social behaviours in 190 laboratory Beagles. Bellamy *et al.* (2018) found associations between the DRD2 gene and fearful behavioural responses in 158 Havanese dogs using owner questionnaires and direct behaviour observations. Significant associations were also identified with the DRD2 gene and noise reactivity in Collie dogs (smooth and rough varieties; n=94) and Irish Soft-Coated Wheaten Terriers (n=44). Results highlight differences between breeds, with a need for breed to be controlled for in genetic evaluation. However caution should be used as the sample size was small. Takeuchi *et al.* (2009a) reported a significant association between activity levels and polymorphisms of T471C in the SLC1A2 gene and the G216A in the COMT gene in a population of 81 USA guide dogs. The methods used by Takeuchi *et al.* (2009a) utilised free text responses of guide dog trainers with reference to specific and often subjective words such as 'nervous' and 'sensitive'. Inconsistency and potential inaccuracy in behavioural data could have impacted the results.

Behavioural traits are complex, and few individual genes have been directly linked to a single behavioural trait (Lewis, 2015). There is evidence to support the contention that behavioural traits are often influenced by multiple genetic and non-genetic factors. By understanding the phenotypic behaviour displayed, reviewing pedigrees and linking to genetic information, it will be possible to understand more about how genetics influence behaviour in the dog.

1.3.5 Guide dog breeding programmes

Guide dog schools have developed breeding programmes to produce dogs suited specifically for the guiding role. As early as 1983, Goddard and Beilharz reported that the Royal Guide Dogs for the Blind Association in Australia were breeding dogs to genetically improve and select those suitable for the guiding role both for health and behaviour. The Seeing Eye (USA) began a guide dog breeding programme in the 1940s (Ennik *et al.*, 2006) and GDUK began their breeding programme in 1960 (Dogs, 2020). Guide dog schools heavily invest in the selection, breeding and upbringing of guide dogs (Ennik *et al.*, 2006; Leighton, 2009), and require temperament information to support breeding for behaviour (Whiteside, 2017). Organisations undertake standardised socialisation (Vaterlaws-Whiteside and Hartmann, 2017), puppy raising and training programmes (Guide Dogs, 2020a) as the early life experiences of puppies have been linked to future behavioural development and behaviours shown in adulthood (Fox and Spencer, 1969; Battaglia, 2009; Foyer *et al.*, 2014). However, the selection for behaviour can lead to the introduction of indirect and unintended traits and a trade-off between different behaviour traits if not monitored appropriately (Goddard and Beilharz, 1983; Asp *et al.*, 2015), so it is important to measure a range of behavioural traits.

GDUK undertakes extensive review of health and behaviour for dogs, siblings and progeny within the breeding programme. The genetic structure of a population can be reviewed using accurate pedigree records (Cole *et al.*, 2004). GDUK has an extensive record of pedigree information for dogs since the breeding programme began in the 1960s, although electronic data collection started later (date unknown but thought to be around 1980's). GDUK have undertaken numerous reviews into the health and behaviour of their dogs over the past decade (Evans, 2014; Harvey, 2014; Asher *et al.*, 2016; Caron-Lormier *et al.*, 2016; Blythe, 2020). Health screening and the use of health EBVs aid in selection for breeding stock and mating decisions.

GDUK previously reviewed behavioural information collected for dogs using subjective withdrawal reasons and observations known as Canine Assessment Summaries (CAS). CAS reports were used by GDUK for many years to record training and behavioural information on dogs throughout puppy raising and training. A review of CAS found the assessments to be unreliable due to subjective labels and terminology used (Asher *et al.*, 2016). However, despite these shortcomings the breeding programme reported using the information successfully in the past and placed alerts on specific behaviours shown in breeding stock progeny (M. Bottomley 2020, *personal communication*, 20 July 2020).

1.3.5.1 Guide dog heritability

The heritability of various health and behaviour traits has been reported for guide dog populations (Goddard and Beilharz, 1983; Schiefelbein, 2012; Evans, 2014; Evans *et al.*, 2015). In addition, EBVs are used by guide dog schools around the world to improve breeding programmes and to identify the next generation of breeding stock (Leighton, 2009). Evans (2014) reviewed the genetic structure of GDUK between 1995 and 2012, focussing on health and behaviour. Evans (2014) examined the heritability of CAS reports for pure and cross breeds, noting the large assessor effect for many behavioural reports and low heritabilities for most of the behaviours assessed. Therefore, implementation of behavioural EBVs have been unsuccessful in GDUK due to the subjective nature of the phenotype data available. Other guide dog schools have reported using measures such as the Behaviour Checklist (BCL) to create behaviour EBVs (Russenberger, 2019) although no published data are currently available to verify the results.

The suggested heritabilities of behavioural traits in guide dog populations have varied depending on the population and traits measured. Bartlett (1976) reported heritability estimates to range from 0.03 for willingness (defined as willingness to do what the dog's master asks of it, regardless of distractions) to 0.25 for ear sensitivity (defined as how loud a vocal

correction the dog required) in a population of 1700 USA guide dogs. These results were reported in a review of canine behavioural genetics; the original PhD is not available online. Therefore, the methods used are unknown, but it is clear from the terminology used to define traits that the methods contained subjective measures. Goddard and Beilharz (1983) described the heritabilities of 17 behavioural factors scored by training staff in a population of 249 Australian Labrador Retriever guide dogs. Heritability estimates ranged from 0 to 0.58. Of the factors measured, nervousness showed the highest heritability (0.58) and was significantly impacted by sire, accounting for 19% of the variation in scores. Schiefelbein (2012) reported heritabilities to range from 0.01 to 0.46 for C-BARQ traits in Labrador Retrievers, Golden Retrievers and German Shepherd Dogs from two guide dog schools in the USA between 2003 and 2010. The majority of heritabilities were below 0.1. Data were gathered by The Seeing Eye (TSE) and Guiding Eyes for the Blind (GEB) for over 3000 dogs at six and 12 months of age. Differences were found between the two schools including for Trainability heritability for Labrador Retrievers at 12 months of age (0.46 TSE and 0.01 GEB) and Stranger-directed aggression heritability at six months (0.13 TSE and 0.01 GEB). Despite many similarities between these organisations, these results confirm the importance of not grouping populations due to potential variations in training practices, qualification criteria and breeding animals.

Heritabilities were reported by Evans *et al.* (2015) for 11 components of a puppy profiling assessment (PPA) that measured behavioural responses to novel stimuli. Over 2000 UK guide dog puppies aged six to seven weeks and were included in analyses and heritability estimates ranged from 0.09 for 'response to assessor following moving object' to 0.25 for the 'following response'. Only Labradors, Golden Retrievers and crosses between these breeds were included in a crossbreed model. Breeds were not separated due to the close relationship between the two breeds, but it would have been interesting for the study to report differences by breed. Overall, results from guide dog populations highlight that even within such

a distinct group (i.e. bred for the same purpose and trained for similar tasks), there are differences between individual populations. Many guide dog schools have developed their own breeding lines over years and have differing methods of training and socialisation, but it is also common for guide dog schools to collaborate for breeding and consequently there are shared genetics between the populations. This highlights the need for heritability estimates to be calculated for different populations and breeds of guide dogs and the need to continually monitor these results. If guide dog schools were to use common language when measuring behaviour/temperament, there may be the opportunity for further collaborative working and improvements internationally between guide dog schools.

The mechanism for measuring and recording traits also needs to be considered when comparing heritabilities between guide dog populations. Smaller populations may show lower heritability than larger populations due to less genetic difference between individuals (Falconer and Mackay, 1989). As heritability is a measure of the genetic variation within a population, smaller populations can often contain a more restricted set of genomes. As such, heritability studies should utilise data from large populations with adequate genetic diversity to ensure genomes are well represented in studies. It is also important to consider the management of a large working dog breeding programme and the ability to obtain data. Working dog programmes, such as guide dogs, often remove dogs that display undesirable behaviours, which can impact the data available for heritability estimates (Goddard and Beilharz, 1983). Furthermore, obtaining behavioural data requires collecting information from many trainers and volunteers with ranging expertise, which could further impact the data reliability and bias.

GDUK currently uses EBVs for health but has previously been unable to use accurate behavioural EBVs. Since the work undertaken by Evans (2014), GDUK has reviewed and improved measures of behavioural monitoring (including implementation of a puppy raising questionnaire),

which should improve information provided to the breeding programme. GDUK could benefit from exploring the use of behavioural EBVs should behavioural traits be heritable.

1.4 Conclusion

The importance of behaviour on the outcome of guide dogs has been widely reported within the literature, with studies aiming to understand the factors influencing success. Due to around 75% of dogs that are unsuccessful in guide dog programmes not qualifying due to behavioural reasons, understanding the ability to influence behaviour by breeding would be of great benefit to save costs, resources and maximise dog welfare. In addition, organisations could breed less dogs, should more of those bred qualify as guide dogs. For many years, GDUK have aimed to understand the genetic influences on behaviour through different projects. Previous analyses have reported limitations in the behavioural phenotype due to the presence of subjective and biased data. This thesis will bring together and build on the previous work undertaken at GDUK, with an emphasis on reviewing the use of terminology when describing behaviour in dogs.

The thesis will aim to understand the genetic influences on guide dog behaviour in a population of puppies ranging from the early emerging behaviour at around seven weeks of age, through to post-juvenile dogs at 12 months of age, with objectives to: 1) determine differences in heritabilities at different ages, 2) compare heritabilities for data collected by questionnaires and behavioural assessments; and 3) review differences in pedigree-based and genomic-based heritabilities.

Chapter 2 will review the suitability of a published behavioural questionnaire (the r-PWQ; Hunt *et al.*, 2020) in puppies at five, eight and twelve months of age collected by Guide Dogs UK for business as usual, with adjustments made the questionnaire component structure to fit the

current population. Chapter 3 will undertake pedigree-based heritability estimates for Labradors, Golden Retrievers and crosses of the two breeds utilising data from Chapter 2. Chapter 4 will present behavioural data collected from a practical assessment in a cohort of puppies around seven weeks of age, where pedigree-based heritabilities will be explored. Chapter 5 will assess inter and intra-rater reliability and determine whether machine learning could be used to improve reliability. Chapter 6 will determine genomic-based heritabilities for 96 dogs using data from Chapter 2 with comparisons made to pedigree-based heritability estimates.

The results will be relevant to Guide Dogs in the UK and other guide and assistance dog organisations internationally. Additionally, the findings will increase the understanding of dog behavioural genetics for the wider dog population.

Chapter 2. Refining a puppy raiser questionnaire from the r-PWQ to r-PWQ2

2.1 Introduction

Monitoring the behaviour of guide dog puppies as they develop through the puppy and juvenile stages is important to ensure dogs are selected appropriately for the working role. A common method for collecting behavioural data for a large and geographically wide-spread population, such as for pet dogs and working dogs, is through questionnaires. By asking the owner of the dog to complete a group of questions, data can be gathered from the individual that typically knows the most about a dog's usual behaviour (Serpell and Hsu, 2001). Identifying behavioural problems early increases the potential for modification and reduces the time for behaviour escalation. Furthermore, puppy welfare is maximised as the puppy will receive appropriate training or will be withdrawn if deemed unsuitable to continue on the guide dog programme.

Many guide dog schools use questionnaires to measure the behaviour of dogs during puppyhood and before being placed into training to support future decision making (Batt *et al.*, 2009; Harvey *et al.*, 2017; Bray *et al.*, 2021c). Harvey (2014) developed the Puppy Walker Questionnaire (PWQ) specifically for use in guide dog populations due to the importance of certain traits and behaviours. A refined version of the questionnaire (the refined Puppy Walker Questionnaire; r-PWQ; Hunt *et al.*, 2020) is now used by Guide Dogs UK (GDUK). The questionnaire consists of 39 questions that form nine traits considered important to guide dogs (separation; excitability; general anxiety; animal chase; trainability; body sensitivity; attachment; energy; distractibility). Scores have been linked to outcome (whether qualified or withdrawn) in two separate cohorts at GDUK. Firstly, a population of 214 puppies at eight months and 226 puppies at 12 months of age (Harvey, 2014) and more recently in a study of 359 puppies at eight months of age (Hunt *et al.*, 2020). The latter study

also reported the r-PWQ to be more suitable when compared to data collected using the Canine Behavioral Assessment & Research Questionnaire (C-BARQ).

There are advantages and disadvantages when collecting data using questionnaires. Questionnaires enable data to be collected with fewer costs and resources and can capture a general overview of the dog's behaviour in a range of different situations and contexts. However, questionnaires have the potential for high levels of bias as owners can interpret and view their dogs differently dependent on animal bond, experience and interpretation (Bellamy *et al.*, 2018). Owners may be more likely to report behavioural problems in pet dogs when survey responses are kept anonymous (Segurson *et al.*, 2005), however anonymity is not possible within guide dogs, as the progression of individuals needs to be monitored. There is the potential for the withholding of information or under-reporting of behaviours, yet this is mitigated where possible by training and support provided to volunteers. Whilst questionnaire data can provide some issues, collecting data in this way enables large-scale studies to be undertaken at low cost with many participants and enables the ability to review genetic variance in populations (Ilska *et al.*, 2017).

The bias from behavioural questionnaires is hard to limit, however if clear terminology and well defined definitions are used in questions, then interpretation differences can be reduced (Goddard and Beilharz, 1983). It is also important that the reliability and validity of behavioural measures are studied (Taylor and Mills, 2006; Harvey, 2014; Harvey, 2021). The scale used can impact variation within questionnaires, for example, the sensitivity differences between Likert scales and visual analogue scales (Cummins and Gullone, 2000), with visual analogue scales able to detect smaller changes. However, variation between individuals will always be present, and is something that in a large, geographically spread programme is difficult to control. It is also important to check the structure of questions and data for different groups of dogs. Due to changes in

breeding and training practices, behaviours may increase or decrease in prevalence in a population between different cohorts of dogs.

Studies utilising questionnaires commonly conduct a principal component analysis (PCA) for each population to ensure questions are grouped appropriately. The goals of a PCA are to extract the most important information, to compress the size and simplify the data (Abdi and Williams, 2010). A PCA can determine both reliability (the consistency of the data) using internal consistency measures (e.g. Cronbach's alpha) and validity (the accuracy of the reports) of questionnaires where content, construct and criterion validity can be explored (Harvey, 2014). It has been suggested that components extracted from a PCA imply potential construct validity (the extent to which components measure a broader construct) due to extracting correlated items within the questionnaire (Hope *et al.*, 2024). Variations are commonly found between different populations when using questionnaires. Friedrich *et al.* (2019) found some C-BARQ questions were not appropriate or did not show enough variation for inclusion in a population of German Shepherd Dogs, so a set of questions were removed. It is essential to ensure that data for a population are treated appropriately for further analyses rather than assumed to be the same as other studies. This chapter aims to review the r-PWQ for a new population of guide dog puppies and determine the most appropriate groupings for further analyses.

2.2 Aims and Objectives

2.2.1 Research question

Is the structure of the r-PWQ questionnaire the same for this population as the one it was developed for?

2.2.2 Aim

To review behavioural data collected at five, eight and 12 months of age using a validated behavioural questionnaire and undertake a PCA to confirm question groupings and create component scores.

2.2.3 Objectives

- To collect behavioural questionnaires from guide dog puppy raisers for puppies of five, eight and 12 months of age.
- To undertake a PCA for questionnaire data at each time point and review the internal consistency of the groupings.
- To calculate component scores based on the outcomes of the PCA.
- To review the impact of additional factors on component scores and the distribution of scores for dogs in the population.

2.3 **Materials and methods**

2.3.1 Study design

A mixture of both retrospective and prospective methods were used to collect behavioural questionnaires for GDUK puppies born over four years (2016-2019) and that were subsequently placed into puppy raising. Data were gathered for puppies at five, eight and 12 months of age using a validated behaviour questionnaire; the refined Puppy Walker Questionnaire, the r-PWQ (Hunt *et al.*, 2020). The five month questionnaire will be referred to as '5M', the eight month questionnaire as '8M' and '12M' refers to the 12 month questionnaire. Data were collected from puppy raisers (volunteers who raise guide dog puppies). Additional data were collected including the puppy raiser ID, details of the puppy raiser home environment, undesirable behaviours shown by the puppy, puppy training location and Puppy Development Advisor (staff members that support puppy raisers). The r-PWQ questions originally formed nine components (Separation, Excitability, Attachment and attention seeking,

Energy Level, General Anxiety, Animal Chase, Body Sensitivity, Distractibility, Trainability). The present study reviewed the internal consistency of the component structures from the r-PWQ and whether using the questionnaire in the existing format was suitable for the current population. Findings suggested the questionnaire structure should be updated for this population, so a PCA was undertaken and the r-PWQ2 developed. Correlations between r-PWQ and r-PWQ2 groupings were compared.

Ethical approval was granted by the University of Nottingham Veterinary School's Committee for Animal Research and Ethics (reference 2798 190814).

The flow diagram below (Figure 2.1) outlines the development of the questionnaire used within the remainder of this thesis, known as the refined Puppy Walker Questionnaire version 2 (r-PWQ2).

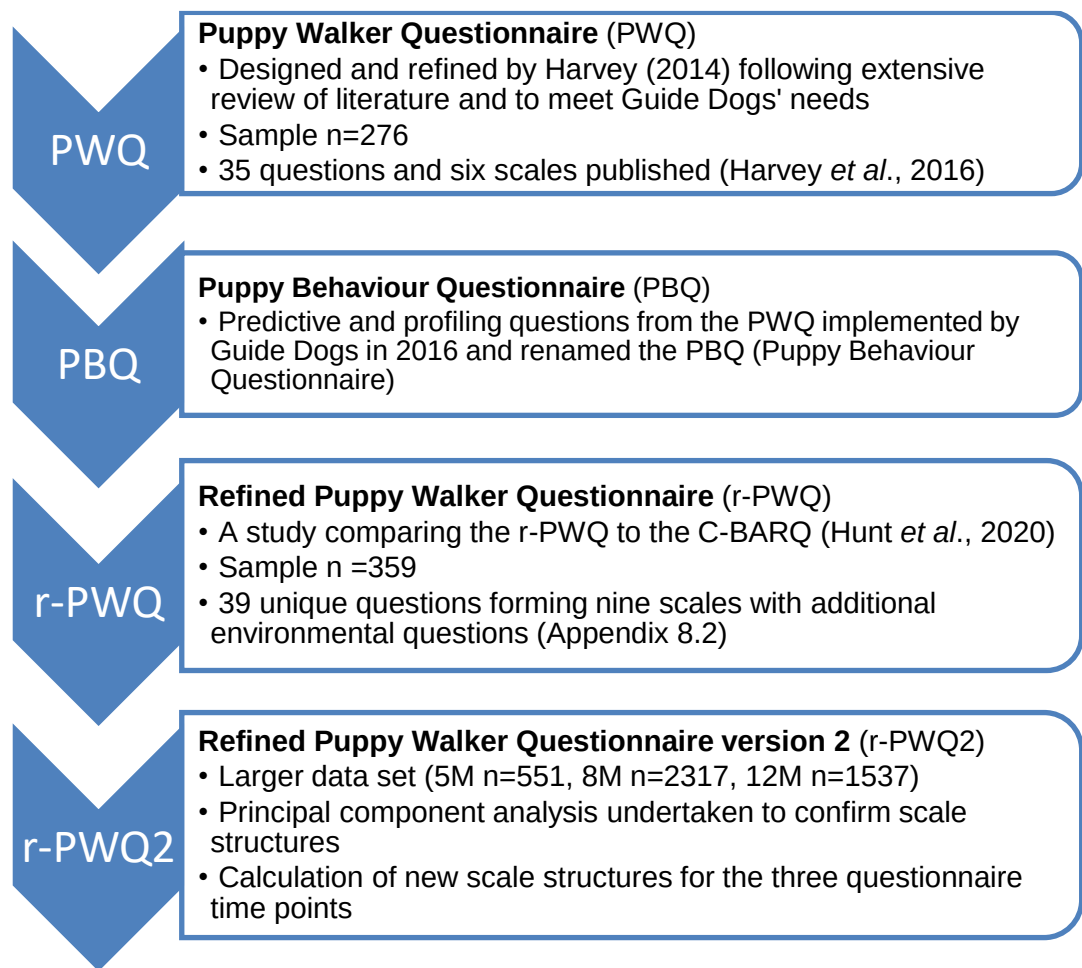


Figure 2.1 Development of the r-PWQ2. The image outlines the four main steps for development and implementation at GDUK.

2.3.2 Study setting

GDUK breed puppies through an internal breeding programme, except for a small number of puppies each year that are sourced externally or from other guide dog schools to maintain genetic diversity within the programme. Puppies are whelped and reared in volunteer family homes or at the Guide Dogs National Centre (GDNC) in a purpose-built kennel facility. The puppies remain with the dam (unless health complications prevent this) until seven weeks of age when they enter the puppy block at the GDNC. Puppies remain at the GDNC for around a week when aged between seven and eight weeks old, and during this time will undergo a health check, microchipping, vaccinations and socialisation activities. Around eight weeks of age puppies travel to new volunteer homes and

enter 'Puppy Raising', where they are socialised in a wide variety of environments similar to those they would experience as a working guide dog and learn basic obedience. Puppies are supervised by staff members known as Puppy Development Advisors (PDA) for the duration of puppy raising. The PDA will monitor the puppies' progress and support the puppy raiser with any issues and appropriate training techniques. Puppies remain in puppy raising until approximately 14 months of age, although the time does vary dependent on the dog's health, behaviour and resources available. Following puppy raising dogs either enter training, at a Guide Dog Training School or Community Team, or if selected as a potential breeding dog, are transferred to the GDNC for a breeding assessment. During training, dogs are either housed in volunteer boarder homes or in kennels and are trained by a mixture of Guide Dog Trainers and Guide Dog Mobility Specialists. Dogs receive reward-based training (involving the addition of praise, food and treats) and are trained through a standardised training approach in two stages (early and advanced training). During early training dogs are taught the essential guiding principles, such as stopping at kerbs and straight-line walking. Dogs then progress to advanced training where they are taught more advanced guiding work in complex environments including traffic work, and a potential match with a client is considered. When a client has been selected and the dog is at a suitable level, both the client and dog enter a stage called 'On Class'. During this stage the dog and client work together on several selected routes before qualifying as a partnership. For breeding dogs, specific litters (sometimes all the litter, and sometimes just males or females) are marked as potential breeding dogs. These dogs are followed through puppy raising and their health and behavioural information is evaluated by the breeding department at set review points. At around 12 months of age, dogs that are still considered suitable for potential breeding dogs are transferred to the GDNC for assessment by the Breeding Programme team. Dogs not suited for the breeding role will return to the guide dog pathway if not accepted. The flow chart below (Figure 2.2) illustrates the dog journey through breeding and training pathways. During

any stage, dogs can be withdrawn for health and/or behavioural reasons that would not be suited to the guiding/breeding role.

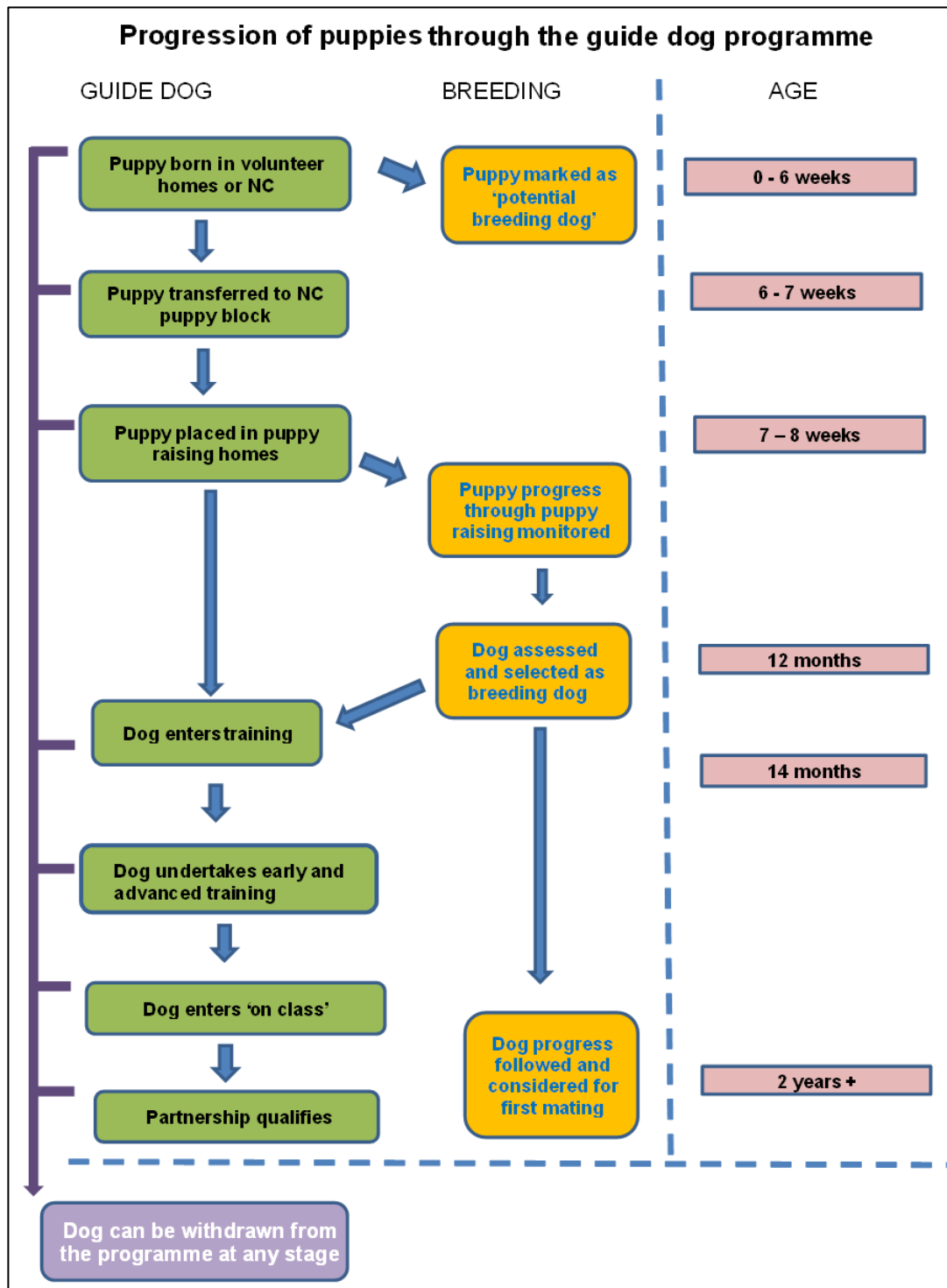


Figure 2.2 Flow diagram of GDUK journey from birth to qualification.

All information on the dogs and volunteers is recorded on GDUK electronic database (Guide Dogs Interactive Replacement, GDIR). All dogs have an individual record, with a six-digit ID as a unique identifier. The ID can be used to search data of interest. Information recorded includes key movement dates, location, staff responsible for the dog, puppy raiser details, health records, behavioural information and qualification/withdrawal information.

2.3.3 Questionnaires

As part of GDUK puppy raising scheme, puppy raisers are required to complete questionnaires on dog behaviour as part of the volunteer role and therefore consent is not collected or required for each questionnaire, or to contact the puppy raiser for information on the puppy. Subjects were GDUK puppies born between 07/08/2016 and 18/07/2019 that were placed into puppy raising. Puppy raisers were emailed a link to an online questionnaire two weeks prior to a puppy turning five, eight and 12 months of age and the questionnaire remained open for four weeks (Appendix 8.1). Puppy raisers with no email address were not contacted during the study and puppies withdrawn from the programme during puppy raising were not sent questionnaires following removal. Puppy raisers opted whether to complete the questionnaire for the puppy in their care. The questionnaire was produced in SurveyGizmo (www.surveygizmo.eu) and contained all 39 items from the r-PWQ along with additional questions to gather information on the puppy raiser's home environment (Appendix 8.2). Following completion of the questionnaire, results were automatically exported into GDIR except for 321 responses where data were downloaded directly from SurveyGizmo as the data interface to export results to GDIR was established during the study.

2.3.4 Variables and data sources

2.3.4.1 *Puppy and puppy raiser information*

Puppy information including date of birth, sex, breed, dam ID, dam name, sire ID, sire name and litter ID were collected from GDIR. Puppy raiser ID, PDA, puppy raising region and team were also obtained from GDIR.

2.3.4.2 *Behavioural questionnaires*

A prospective and retrospective cohort method was used in this study to gather behavioural data on Guide Dogs' puppies at five, eight and 12 months of age using a validated behavioural questionnaire (r-PWQ). Data collected at five months was collected prospectively and eight and 12 month data collected retrospectively. Data were collected from puppy raisers and information taken from GDIR.

2.3.5 Bias and study size

Study size was determined by the number of puppies produced by GDUK and placed into puppy raising and the number of completed questionnaires by puppy raisers. The study size was reduced by puppy raisers not having email addresses (therefore not receiving the questionnaire) and by puppies being removed from the programme during puppy raising. The exact number of puppy raisers without an email address is unknown, but as volunteers apply for the role online via an email address, and trainings material are provided online, the numbers are expected to be small. Puppies removed (withdrawn) from puppy raising before entering training are around 35% (Moxon and Fellows, 2022), but the age at withdrawal will vary depending on the health or behaviour reason. There is a tendency to process withdrawals around 12 months before entry to training, so withdrawal had a limited impact on the sample size.

The sample size was smaller than originally planned as data collection ceased due to the Covid-19 pandemic and the impact on the phenotype data collected during this period. Data collected after March 2020 was not included in the study due to the impact of reduced socialisation on puppies, therefore not experiencing “normal” puppy raising, and due to the impact of the pandemic on puppy raiser emotions and positivity and thus potential for negative bias in questionnaire responses. Bias in puppy raiser opinions of their puppy could also be present within the sample. Puppy raisers may not want to report negative or undesirable behaviours seen as they may feel it would reflect on them as a puppy raiser. As responses are not anonymous, this could result in the withholding of information, although all puppy raisers are encouraged to be as honest as possible in the introduction to the questionnaire.

2.3.6 Quantitative variables and statistical methods.

2.3.6.1 *Checking suitability of r-PWQ*

The published r-PWQ (Hunt *et al.*, 2020) was checked for suitability in the current population. The internal reliability of the behaviour components were checked using Cronbach’s alpha (Table 2.1).

Table 2.1 Internal reliability of 5M, 8M and 12M r-PWQ traits

r-PWQ trait name	5M Cronbach's α	8M Cronbach's α	12M Cronbach's α
Separation-Related Behaviour	0.82	0.74	0.79
Excitability	0.72	0.73	0.72
General Anxiety	0.85	0.83	0.83
Animal Chase	0.75	0.77	0.79
Distractibility	0.84	0.82	0.80
Trainability	0.61	0.62	0.62
Body Sensitivity	0.78	0.76	0.77
Attachment and attention seeking	0.62	0.60	0.60
Energy Level	0.85	0.85	0.81

Results suggested that two of the nine behavioural traits at all three time points (Excitability and Attachment and attention seeking) displayed a lower than acceptable Cronbach's alpha, described as being below 0.7 (Tavakol and Dennick, 2011). The lowest Cronbach's alpha was found for Attachment and attention seeking. The aim of the present study was to investigate genetic influences on behaviour, and it was important to ensure the most appropriate component structure was used during analyses. A revised version of the r-PWQ was developed using a principal component analysis to fit the current population.

2.3.6.2 Principal component analyses

A principal component analysis (PCA) was undertaken using raw questionnaire scores to determine the structure of the questionnaire grouping and to determine the proportion of variation in the data set accounted for by components. Results from a PCA are uncorrelated components (or eigenvectors) that capture variance in the data set (Hope *et al.*, 2024). It is recommended that during a PCA analysis, the outputs are appropriately extracted, rotated and interpreted (Tavakol and Wetzel, 2020) to ensure validity. Rotations aid with interpretation of results and loadings are correlations between a variable and a component (Abdi and Williams, 2010). The varimax rotation is the most commonly used rotation where components represent a small number of variables (Abdi and Williams, 2010).

A separate component analysis was undertaken for each data set (5M, 8M, 12M). The 'Is self-controlled and calm' was asked twice in the r-PWQ to determine the internal consistency for individuals completing the questionnaire and because the question featured in two components during development. The agreement in scores between the first and second time the question was asked was tested using Spearman's rank as data were not normally distributed. Mukaka (2012) reports correlation coefficients of over 0.9 to be very high, 0.7 to 0.9 to be high, 0.50 to 0.70 to be moderate, 0.3 to 0.5 and any correlation below 0.3 as negligible. The

score for the first time the question was asked was chosen to correlate to GDUK practices (as the question was only asked once to reduce question numbers).

The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity were used to determine whether a PCA was appropriate for use in each data set. After the components were determined, a rotation was used to facilitate interpretation (Abdi and Williams, 2010). The PCA utilised a correlation matrix, with varimax rotation. A PCA was run using an Oblim rotation, but few factor correlations values exceeded 0.32 (Tabaknick and Fidell, 2007) and therefore the rotation was not suitable for the data. Items with extraction communalities below 0.4 were removed and the PCA rerun until all communalities were above 0.4. Only items with loadings of >0.4 were included in the scales.

Internal reliability of the scales were tested using Cronbach's alpha statistics. A Cronbach's alpha of greater than 0.70 was considered to indicate good internal reliability. However, as all scales had fewer than 10 items, a Cronbach's alpha of greater than 0.5 was considered acceptable (Van Griethuijsen *et al.*, 2015; Taber, 2018). All statistics for PCAs were undertaken using SPSS version 22 (SPSS Inc., Chicago, IL). Data were presented as mean $\pm$ standard error.

2.3.6.3 Calculating component scores

Questions were scored using a sliding bar visual analogue scale from 0-100, where 0 indicated the behaviour had never been seen and 100 the behaviour was seen all the time. Component scores were calculated using the same method as previously described for the PWQ (Harvey, 2014) and the r-PWQ (Hunt *et al.*, 2020). The mean of all questions within each component was taken to create a component score. Only complete questionnaires were used to create component scores, therefore eight incomplete 8M questionnaire results from the PCA analysis were

removed. Scores for the 'Focus' component at 5M (Component 10) and 8M (Component 8) were reversed so 100 equated to an undesirable behaviour. Therefore, for all components 0 was a more desirable score, however arguably there was no desirable score for 'Energy' as puppies show a range of energy levels. Components are referred to as C followed by the number, e.g. Component 1 is 'C1'.

Histograms of individual component scores at all three time points were produced to determine the score distribution (Appendix 8.4). Minimum, maximum, mean and standard deviations for all components were calculated. The age at completion (months) was calculated for all questionnaire responses. Individual Chi-squared tests with a sequential Bonferroni correction were used to test the difference between breed numbers for 5M, 8M and 12M samples. Breeds with smaller numbers were grouped for comparisons (Curly Coated Retriever, Curly Coated Retriever x Labrador, Labrador x Standard Poodle and Standard Poodle x Labrador were grouped, Golden Retriever x Golden Retriever* were grouped with Golden Retriever x Labrador* and Labrador x Golden Retriever* were grouped with Labrador x Labrador*). Breeds were abbreviated so Golden Retrievers are reported as 'GR', Labradors as 'Lab', German Shepherd Dogs as 'GSD', Standard Poodles as 'SP' and Curly Coated Retrievers as 'CCR'. Breeds are reported sire first with an x indicating a cross breed and a * indicating the dam was also a crossbreed. Chi-squared tests were used to compare the number of males and females with completed responses at each time point. A univariate analysis of variance was used to determine whether breed and sex impacted component scores. Analyses were undertaken in SPSS version 22 (SPSS Inc., Chicago, IL).

2.3.6.4 Additional factors

Additional data were collected on the home setting (number of adults in the home, children, presence of another dog, animals (other than dogs), the location of the house (rural, suburban, urban), the age (closest month)

at neutering, whether puppy raisers had attended puppy classes and if the puppy had been taken into the workplace.

Litter and PDA were coded into numerical format (from one to number of subjects). Puppy raiser experience was grouped into four groups (raised under five puppies, raised five to nine puppies, raised 10 to 19 puppies and raised 20 or more puppies). The number of adults in the home was recorded and the presence of five and over grouped together. The presence of children, other dogs, animals (other than dogs) were coded as 'yes' or 'no' rather than the number of each factor. House location was recoded as either urban, suburban, or rural and attendance at a puppy class and visiting the workplace were recorded as 'yes' or 'no'.

The differences in additional information variables between the three time points were tested using Chi-squared tests. Count data for the different variables were compared with the number of missing values not included. For number of adults in the home, the presence of four adults and 5+ adults were grouped. Analyses were undertaken using an online Chi-squared calculator (<http://quantpsy.org/chisq/chisq.htm>).

Neutering dates were extracted from GDIR health records. Any neutering date that occurred over 18 months of age was manually checked to ensure the dates were correct and any dogs with discrepancies were removed from the data set. Neutering dates for dogs selected as breeding stock at any point were removed as this would impact age at neutering. Some dogs were selected as breeding dogs and then subsequently either returned to training or withdrawn and then neutering dates were delayed. The age at neuter (months) and days of age at neuter were calculated. Whether the dog was neutered by the time of questionnaire expiry was determined for eight (neutered by 8.5 months of age) and 12 month (neutered by 12.5 months of age) data sets so two groups were available.

A general linear model was run (GLM) for each behaviour component to assess the impact of different factors on the behaviour scores. Standardised residuals were checked for normality and if data were not normally distributed, a generalized linear model (GLZ; multiple linear regression) was run. All models included the factors of sex, litter, PDA, Puppy raiser experience group, number of adults, presence of children, presence of dogs, presence of animals, house location, attendance at puppy class and visiting the workplace. Year of birth was determined from the date of birth and included in 8M and 12M models. The age at neuter (in months) was included in 8M and 12M models (the model does not account for whether the dog was neutered at the time of questionnaire completion). It was not possible to include breed in the model due to the number of combinations available for the different factors. A separate generalized linear model (GLZ) was run for each component to determine breed and sex effects.

2.3.6.5 Comparing the r-PWQ and r-PWQ2

A comparison of the relating r-PWQ and r-PWQ2 components were compared at all three time points (5M, 8M and 12M). A Pearson's correlation coefficient was undertaken using SPSS version 29 (SPSS Inc., Chicago, IL).

2.4 Results

2.4.1 Participants

The study population for questionnaire responses included 2892 dogs (1478M/1414F) born between 07/08/2016 and 18/07/2019 and subsequently placed into GDUK's puppy raising programme between around eight weeks of age. Dogs were bred and reared by GDUK except for 23 dogs that were bred and reared (to eight weeks of age) by external breeders. Dogs were from 14 breeds (Table 2.2) and were produced by 112 sires (one external sire) and 350 dams (one external dam) from 522

litters. 5M data were collected for 551 dogs (19.1%), 8M data for 2317 dogs (80.1%) and 12M data for 1537 dogs (53.1%), with a total of 4405 questionnaires collected. The number of questionnaires varied for individual dogs with one questionnaire collected for 1380 dogs (47.7%), two questionnaires collected for 1511 dogs (52.2%) and one dog had all three questionnaires collected.

Table 2.2 Breeds/crossbreeds represented in the r-PWQ2 study population. Number and percentage of breeds for behavioural questionnaires collected at five (5M), eight (8M) and 12 (12M) of age.

Breed	Number (%) in population	Number (%) at 5M	Number (%) at 8M	Number (%) at 12M
GR x Lab	704 (24.3%)	114 (20.7%)	542 (23.4%)	428 (27.9%)
Lab	695 (24.0%)	156 (28.3%)	584 (25.2%)	301 (19.6%)
Lab x GR	332 (11.5%)	41 (7.4%)	251 (10.8%)	224 (14.6%)
GR	316 (10.9%)	100 (18.2%)	249 (10.7%)	126 (8.2%)
Lab x GR ^{*3}	258 (8.9%)	16 (2.9%)	225 (9.7%)	168 (10.9%)
GR x GR ^{*2}	239 (8.3%)	54 (9.8%)	193 (8.3%)	105 (6.6%)
GSD	146 (5.0%)	46 (8.4%)	113 (4.9%)	57 (3.7%)
GR x GSD	107 (3.7%)	8 (1.5%)	84 (3.6%)	70 (4.6%)
GR x Lab ^{*2}	33 (1.1%)	6 (1.1%)	28 (1.2%)	19 (1.2%)
SP x Lab ^{*4}	21 (0.7%)	5 (0.9%)	15 (0.6%)	12 (0.8%)
Lab x Lab ^{*3}	19 (0.7%)	5 (0.9%)	12 (0.5%)	13 (0.9%)
Lab x SP ^{*4}	15 (0.5%)	0 (0.0%)	14 (0.6%)	9 (0.6%)
CCR x Lab ^{*1}	5 (0.2%)	0 (0.0%)	5 (0.2%)	4 (0.3%)
CCR ^{*1}	2 (0.1%)	0 (0.0%)	2 (0.1%)	1 (0.1)

*Breeds and crossbreeds are listed as sire first and ordered from largest to smallest overall representation. Superscript numbers (^{1,2,3,4}) indicate the breeds that were grouped together for analyses. Breeds are abbreviated to Lab (Labrador), GR (Golden Retriever), GSD (German Shepherd Dog), SP (Standard Poodle) and CCR (Curly Coated Retriever).

The number of breeds represented for each time point were compared and corrected for using Chi-squared with sequential Bonferroni for multiple testing. When comparing the 5M sample to the 8M sample there were significant differences for 12 of the breed combinations. Lab x GR* and Lab x Lab* were grouped together due to smaller numbers. There were differences between this breed group and the number of GR ($\chi^2=39.11$, D.F.=1, $P<0.001$), GSD ($\chi^2=35.53$, D.F.=1, $P<0.001$), Lab ($\chi^2=21.96$,

D.F.=1, $P<0.001$), GR x Lab ($\chi^2=12.56$, D.F.=1, $P<0.001$) and GR x GR* grouped with GR x Lab* ($\chi^2=18.39$, D.F.=1, $P<0.001$). There were differences in the numbers of GSD when compared to Lab x GR ($\chi^2=14.66$, D.F.=1, $P<0.001$), GR x Lab ($\chi^2=10.83$, D.F.=1, $P<0.001$) and GR x GSD ($\chi^2=14.13$, D.F.=1, $P<0.001$), and the numbers of GR when compared to Lab x GR ($\chi^2=19.78$, D.F.=1, $P<0.001$), GR x Lab ($\chi^2=17.28$, D.F.=1, $P<0.001$) and GR x GSD ($\chi^2=15.68$, D.F.=1, $P<0.001$).

When comparing the 5M sample to the 12M sample there were significant differences for 17 of the breed combinations. There were differences in the numbers of Lab compared to GR x GSD ($\chi^2=17.87$, D.F.=1, $P<0.001$), GR x Lab ($\chi^2=21.58$, D.F.=1, $P<0.001$), Lab x GR ($\chi^2=29.45$, D.F.=1, $P<0.001$) and Lab x GR* grouped with Lab x Lab* ($\chi^2=40.19$, D.F.=1, $P<0.001$). There were differences in the number of GR x GR* grouped with GR x Lab* compared to GR x GSD ($\chi^2=14.24$, D.F.=1, $P<0.001$), GR x Lab ($\chi^2=10.10$, D.F.=1, $P<0.001$), Lab x GR ($\chi^2=18.29$, D.F.=1, $P<0.001$) and Lab x GR* grouped with Lab x Lab* ($\chi^2=28.65$, D.F.=1, $P<0.001$). There were differences in the number of GSD compared to GR x GSD ($\chi^2=25.10$, D.F.=1, $P<0.001$), GR x Lab ($\chi^2=25.90$, D.F.=1, $P<0.001$), Lab x GR ($\chi^2=35.01$, D.F.=1, $P<0.001$) and Lab x GR* grouped with Lab x Lab* ($\chi^2=46.72$, D.F.=1, $P<0.001$). GR numbers were different to GR x GSD ($\chi^2=29.25$, D.F.=1, $P<0.001$), Lab x GR ($\chi^2=49.34$, D.F.=1, $P<0.001$) and to Lab x GR* grouped with Lab x Lab* ($\chi^2=60.28$, D.F.=1, $P<0.001$). There were differences in GR x Lab compared to Lab x GR* grouped with Lab x Lab* ($\chi^2=11.1$, D.F.=1, $P<0.001$) and compared to GR ($\chi^2=42.76$, D.F.=1, $P<0.001$).

When comparing the 8M sample to the 12M sample there were significant differences for six breed combinations. There were differences in the number of Lab when compared to Lab x GR* grouped with Lab x Lab* ($\chi^2=10.51$, D.F.=1, $P<0.001$), Lab x GR ($\chi^2=22.54$, D.F.=1, $P<0.001$) and GR x Lab ($\chi^2=20.07$, D.F.=1, $P<0.001$). There were differences in the number of GR x Lab compared to GR ($\chi^2=12.50$, D.F.=1, $P<0.001$) and

GR x GR* grouped with GR x Lab* compared to Lab x GR ($\chi^2=10.29$, D.F.=1, $P<0.001$).

Data were collected for 275 males and 276 females at 5M, 1186 males and 1131 females at 8M and 785 males and 752 females at 12M. There was no difference in sex distribution between the three questionnaire time points ($\chi^2=0.29$, D.F.=2, $P=0.87$).

2.4.1.1 5M participants

A total of 551 completed 5M questionnaires were used for analyses from 275 males and 276 females. Dogs comprised of 11 breeds or crossbreeds (Table 2.2) and were aged 4.81 months (± 0.25) at questionnaire completion. Dogs were from 47 sires and 102 dams and from 102 litters. Questionnaires were completed by 548 individual puppy raisers. Three puppy raisers completed two 5M questionnaires and the remaining 545 completed one questionnaire.

2.4.1.2 8M participants

A total of 2317 completed 8M questionnaires were used for analyses from 1186 males and 1131 females. Dogs comprised of 14 breeds or crossbreeds (Table 2.2) and were aged 7.64 months (± 0.24) at questionnaire completion. Dogs were from 111 sires and 330 dams and from 476 litters. Questionnaires were completed by 1667 individual puppy raisers and one additional puppy raiser was recorded as 'missing' as there was conflicting data on the individual that had completed the questionnaire. Ten puppy raisers completed four questionnaires, 105 completed three questionnaires, 409 completed two questionnaires and the remaining 1143 completed one questionnaire.

2.4.1.3 12M participants

A total of 1537 completed 12M questionnaires were used for analyses from 785 males and 752 females. Dogs comprised of 14 breeds or crossbreeds (Table 2.2) and were aged 11.68 months (± 0.21) at questionnaire completion. Dogs were from 92 sires and 272 dams and from 342 litters. Questionnaires were completed by 1263 individual puppy raisers. One puppy raiser completed five questionnaires, 24 completed three questionnaires, 222 completed two questionnaires and the remaining 1016 puppy raisers completed one questionnaire.

2.4.2 Principal component analyses (PCA)

The agreement in scores between the first and second time the 'Is self-controlled and calm' question was compared for all time points. There was a significant correlation between the first and second time the question was asked at 5M ($r_s=0.60$, $P<0.001$, $n=551$), 8M ($r_s=0.62$, $P<0.001$, $n=2317$), and 12M ($r_s=0.63$, $P<0.001$, $n=1537$). There were moderate correlations between the questions at all time points. Results suggested it was not appropriate to take an average of the two scores.

2.4.2.1 5M PCA

Ten components emerged from the PCA with eigenvalues >1 , explaining 64.7% of the total variance in the data set. Eigenvalues are numerical values associated with variation present in the components extracted during a PCA. Two questions were removed as the communalities for the individual questions were below 0.4. Two questions removed were 'Loses its appetite when left, or about to be left' and 'Returns directly to you if startled or frightened'. Communalities were above 0.4 for all remaining questions (Appendix 8.3.1). The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.841 (minimum acceptable KMO is 0.50, Budaev 2010) and Bartlett's Test of Sphericity was $P<0.001$, suggesting a PCA was appropriate for the data (Budaev, 2010). C1 accounted for

21.2% of the variation found within the data set and all other components accounted for below 10.0% of the variation (Table 2.3). Results from the component matrix indicate over 25% of the variation was present in the sample for the first two components and over 50% in the first six components. Full component loadings can be found in Appendix 8.3.1.

Table 2.3 Components extracted from 5M PCA. Total eigenvalue, percentage of variance and cumulative percentage of variance for 10 components.

Component	Total eigenvalue	Percentage of variance (%)	Cumulative percentage (%)
C1	7.8	21.2%	21.2%
C2	2.9	7.8%	28.9%
C3	2.4	6.4%	35.4%
C4	2.0	5.5%	40.9%
C5	1.8	5.1%	46.0%
C6	1.5	4.1%	50.1%
C7	1.3	3.5%	53.6%
C8	1.2	3.5%	57.1%
C9	1.2	3.3%	60.4%
C10	1.1	3.0%	63.4%

Cronbach's alpha scores ranged from 0.47 to 0.85 (Table 2.4). Two components had a Cronbach's alpha below 0.7 (C5=0.65 and C10=0.47). See Table 2.9 for questions included in each component.

Table 2.4 Internal reliability for 5M PCA for the 10 extracted components.

Component	Number of questions	Cronbach's alpha
C1	9	0.83
C2	5	0.84
C3	4	0.85
C4	6	0.74
C5	5	0.65
C6	5	0.80
C7	2	0.85
C8	2	0.78
C9	2	0.78
C10	4	0.47

Following a review of the questions present in the 5M sample and the component names used by Harvey (2014), new component names have been suggested for: C1-Distraction and Human Focus, C4-General

Excitement, C6-Impulsivity and C8-Anticipatory Excitement. The other component names remain the same as those used by Harvey (2014) and Hunt *et al.*, (2020).

2.4.2.2 8M PCA

Ten components emerged from the PCA with eigenvalues >1, explaining 64.1% of the total variance. Six questions were removed as the communalities for the individual questions were below 0.4. These were 'Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left', 'Loses its appetite when left, or about to be left', 'Is hyperactive, restless, has trouble settling down', 'Barks persistently when alarmed or excited', 'Is motivated towards/distracted by food on the ground and/or on tables/shelves' and 'Returns directly to you if startled or frightened'. Communalities were above 0.4 for all remaining questions (Appendix 8.3.2). The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.824 and the Bartlett's Test of Sphericity was $P < 0.001$, suggesting a PCA was appropriate for the data. C1 accounted for 20.0% of the variation within the data set and all other components accounted for below 10.0% of the variation (Table 2.5). Results from the component matrix indicate that over 50% of the variation was present in the first seven components. Full component loadings can be found in Appendix 8.3.2

Table 2.5 Components extracted from 8M PCA. Total eigenvalue, percentage of variance and cumulative percentage of variance for 10 components.

Component	Total eigenvalue	Percentage of variance (%)	Cumulative percentage (%)
C1	6.6	20.0%	20.0%
C2	2.4	7.4%	27.4%
C3	2.2	6.6%	34.0%
C4	1.9	5.8%	39.8%
C5	1.6	5.0%	44.8%
C6	1.5	4.6%	49.5%
C7	1.4	4.1%	53.6%
C8	1.3	3.8%	57.4%
C9	1.1	3.4%	60.8%
C10	1.1	3.3%	64.1%

Cronbach's alpha scores ranged from 0.51 to 0.85 (Table 2.6). Three components had a Cronbach's alpha below 0.7 (C 6=0.58, C8=0.51 and Component 10=0.63). See Table 2.9 for questions included in each component.

Table 2.6 Internal reliability for 8M PCA for the 10 extracted components.

Component	Number of questions	Cronbach's alpha
C1	7	0.84
C2	4	0.83
C3	4	0.76
C4	4	0.71
C5	2	0.82
C6	3	0.58
C7	2	0.85
C8	3	0.51
C9	2	0.76
C10	2	0.63

Following a review of the questions present in the 8M sample and the component names used by Harvey (2014), new component names have been suggested for: C1-Impulsivity, C5-Anticipatory Excitement, C6-Attachment and C10-Attention Seeking. The other component names remain the same as those used by Harvey (2014) and Hunt *et al.*, (2020).

2.4.2.3 12M PCA

Ten components emerged from the PCA with eigenvalues >1, explaining 61.5% of the total variance. Three questions were removed as the communalities for the individual questions were below 0.4. These were 'Loses its appetite when left, or about to be left', 'Is motivated towards/distracted by food on the ground and/or on tables/shelves' and 'Returns directly to you if startled or frightened'. Communalities were above 0.4 for all remaining questions (Appendix 8.3.3). The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was 0.848 and the Bartlett's Test of Sphericity was $P < 0.001$, suggesting a PCA was appropriate for the data. C1 accounted for 20.8% of the variation within the data set and all other components accounted for below 10.0% of the variation (Table 2.7).

Results from the component matrix indicate that over 50% of the variation was present in the first seven components. Full component loadings can be found in Appendix 8.3.3.

Table 2.7 Components extracted from 12M PCA. Total eigenvalue, percentage of variance and cumulative percentage of variance for 10 components.

Component	Total eigenvalue	Percentage of variance (%)	Cumulative percentage (%)
C1	7.5	20.8%	20.8%
C2	2.4	6.7%	27.6%
C3	2.2	6.1%	33.7%
C4	1.9	5.2%	38.9%
C5	1.7	4.8%	43.7%
C6	1.5	4.3%	48.0%
C7	1.3	3.7%	51.7%
C8	1.3	3.7%	55.4%
C9	1.2	3.2%	58.6%
C10	1.1	2.9%	61.5%

Cronbach's alpha scores ranged from 0.34 to 0.83 (Table 2.8). Three components had a Cronbach's alpha below 0.7 (C5=0.64, Component 6=0.66 and C10=0.34). C10 showed very low internal consistency and suggested the questions should not be used as a reliable component. See Table 2.10 for questions included in each component.

Table 2.8 Internal reliability for 12M PCA for the 10 extracted components.

Component	Number of questions	Cronbach's alpha
C1	5	0.80
C2	4	0.83
C3	5	0.82
C4	6	0.76
C5	5	0.64
C6	5	0.66
C7	2	0.86
C8	2	0.81
C9	2	0.77
C10	3	0.34

Following a review of the questions present in the 12M sample and the component names used by Harvey (2014), new component names have been suggested for: C3-Impulsivity, C6-General Excitement, C7-

Anticipatory Excitement and C10-Human Focus. The other component names remain the same as those used by Harvey (2014) and Hunt *et al.*, (2020).

2.4.2.4 Summary of PCA results

All three analyses resulted in the extraction of 10 components following PCA with a varimax rotation accounting for approximately 65% of the variation in all data sets (64.7% at 5M, 64.1% at 8M and 61.5% at 12M). The first component for each PCA accounted for 20% of the variation in all data sets (20.7% at 5M, 20.0% at 8M and 20.8% at 12M). The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy was over the minimum acceptable limit of 0.5 for all three analyses run (0.841 at 5M, 0.824 at 8M and 0.848 at 12M). For all three PCA analyses, two questions were removed 'Loses its appetite when left, or about to be left' and 'Returns directly to you if startled or frightened', with an additional four questions removed at 8M and one more removed at 12M. The internal reliability was acceptable (Cronbach's alpha >0.7) for most components at all three time points with two components at 5M and three components at 8M and 12M below 0.7 (Table 2.9). The 10 components below are not directly comparable between the three time points. See Table 2.9 for questions within each component.

Table 2.9 Summary of component Cronbach's alpha (α) for 10 components collected at five (5M), eight (8M) and 12 (12M) of age.

Component	5M α	8M α	12M α
C1	0.83	0.84	0.80
C2	0.84	0.83	0.83
C3	0.85	0.76	0.82
C4	0.74	0.71	0.76
C5	0.65	0.82	0.64
C6	0.80	0.58	0.66
C7	0.85	0.85	0.86
C8	0.78	0.51	0.81
C9	0.78	0.76	0.77
C10	0.47	0.63	0.34

*NB. the questions asked in each component vary between the time points based on the PCA findings above.

There were some similarities and differences in the structure of the components and questionnaire structure between the three time points (Table 2.10).

Table 2.10 r-PWQ2 component structure. Question number, wording and component structure for five (5M), eight (8M) and 12 (12M) data sets. Colours have been used to illustrate similar components.

Q	r-PWQ question wording	5M	8M	12M
1	Appears restless/agitated or paces when left, or about to be left	Separation	Separation	Separation
2	Whines when left, or about to be left	Separation	Separation	Separation
3	Barks when left, or about to be left	Separation	Separation	Separation
4	Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left	Separation	N/A	Separation
5	Loses its appetite when left, or about to be left	N/A	N/A	N/A
6	Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	Separation	Separation	Separation
7	Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk	Anticipatory Excitement	Anticipatory Excitement	Anticipatory Excitement
8	Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip	Anticipatory Excitement	Anticipatory Excitement	Anticipatory Excitement
9	Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	General Excitement	Impulsivity	General Excitement
10	Is excessive, difficult to control and if it lunges is hard to hold back	General Excitement	Impulsivity	General Excitement
11	Is hyperactive, restless, has trouble settling down	General Excitement	N/A	General Excitement
12	Barks persistently when alarmed or excited	General Excitement	N/A	General Excitement
13	Is self-controlled and calm	General Excitement Focus	Focus	General Excitement
14	Is obviously disturbed by loud or unexpected sounds	Anxiety	Anxiety	Anxiety

15	Is spooked by odd or unexpected things or objects	Anxiety	Anxiety	Anxiety
16	Is anxious or uneasy in new situations	Anxiety	Anxiety	Anxiety
17	Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)	Anxiety	Anxiety	Anxiety
18	Chases birds or squirrels (or would like to)	Impulsivity	Impulsivity	Impulsivity
19	Chases cats (or would like to)	Impulsivity	Impulsivity	Impulsivity
20	Pulls (including lunging) towards unfamiliar dogs	Distraction and Human Focus	Impulsivity	Impulsivity
		General Excitement		
		Impulsivity		
21	Is motivated towards/distracted by food on the ground and/or on tables/shelves	Distraction and Human Focus	N/A	N/A
		Impulsivity		
22	Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	Distraction and Human Focus	Impulsivity	Impulsivity
				Focus
23	Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	Distraction and Human Focus	Impulsivity	Impulsivity
		Impulsivity		Focus
24	Attention can be attracted easily but it loses interest soon	Distraction and Human Focus	Distraction	Focus
25	Will look at you when you talk to it directly in the home environment	Focus	Focus	Human Focus
26	Attention can be easily distracted	Distraction and Human Focus	Distraction	Focus
27	Needs obedience commands repeating to get a response	Distraction and Human Focus	Distraction	Focus

28	Seems like it doesn't listen even if it knows someone is speaking to it	Distraction and Human Focus	Distraction	Focus
		Focus		
29	Stays/waits when instructed to	Focus	Focus	Human Focus
30	Is uneasy with being physically handled/groomed	Body Sensitivity	Body Sensitivity	Body Sensitivity
31	Attempts to move away when you start to groom it	Body Sensitivity	Body Sensitivity	Body Sensitivity
32	Tends to follow you (or other member of household) about the house from room to room	Attachment and Attention Seeking	Attachment	Attachment and Attention Seeking
33	Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	Distraction and Human Focus	Attachment	Attachment and Attention Seeking
		Attachment and Attention Seeking		
34	Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person	Attachment and Attention Seeking	Attention Seeking	Attachment and Attention Seeking
				Human Focus
35	Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal	Attachment and Attention Seeking	Attention Seeking	Attachment and Attention Seeking
36	Displays a strong attachment for one particular member of the household	Attachment and Attention Seeking	Attachment	Attachment and Attention Seeking
37	Returns directly to you if startled or frightened	N/A	N/A	N/A
38	Is active and energetic	Energy	Energy	Energy
39	Is playful	Energy	Energy	Energy

2.4.3 Refined Puppy Walker Questionnaire version 2 (r-PWQ2)

2.4.3.1 *5M descriptive data*

Mean component scores ranged from 8.76 for Separation to 84.90 for Energy; Table 2.11). All components had a minimum score of 0 except for Component 1 at 1.67. Three components had a maximum score of 100 (Separation, Energy and Anticipatory Excitement) and the lowest maximum score was 78.75 for Focus. Scores at 5M suggest that as a population, puppies on average scored highest for Energy levels. After Energy, the highest mean scores were reported for C1 at 39.79 (Distraction and Human Focus) and C6 (Impulsivity) at 38.43. See Appendix 8.4.1 for component distributions.

Table 2.11 5M r-PWQ2 component scores. Minimum, maximum, mean ($\pm$ standard error) for 10 behavioural components.

C	5M r-PWQ	Min	Max	Mean	$\pm$ S.E.
C1	Distraction and Human Focus	1.67	92.11	39.79	0.77
C2	Separation	0.00	100.00	8.76	0.57
C3	Anxiety	0.00	91.00	12.71	0.64
C4	General Excitement	0.00	89.00	25.08	0.70
C5	Attachment and Attention Seeking	0.00	95.80	29.74	0.75
C6	Impulsivity	0.00	99.80	38.43	0.96
C7	Energy	0.00	100.00	84.90	0.71
C8	Anticipatory Excitement	0.00	100.00	9.36	0.68
C9	Body Sensitivity	0.00	94.50	12.75	0.79
C10	Focus	0.00	78.75	21.53	0.60

*A lower score is more desirable and '0' indicates the dog never performs the behaviour, except for C7 where no score is desirable.

One component (Distraction and Human Focus) was significantly impacted by sex and four components (Separation, Anxiety, General Excitement and Focus) were significantly impacted by breed (Table 2.12). For a breakdown of mean component score by breed and sex see Table 8.9 in Appendix 8.5.1. Females had more desirable r-PWQ2 scores in three components, compared to males having more desirable scores in seven components. Favourable r-PWQ2 component scores varied by

breed, with the Lab x GR* and Lab x Lab* (n=21) group having the highest number of desirable scores when compared to other breeds.

Table 2.12 Sex and breed effects for 5M behavioural components, reporting Wald Chi-square, degrees of freedom (df) and P-value.

C	Sex effect			Breed effect		
	Wald Chi-Square	df	P value	Wald Chi-Square	df	P-value
C1	3.92	1	0.05*	14.54	8	0.07
C2	1.20	1	0.27	20.06	8	0.01*
C3	0.71	1	0.40	30.77	8	<0.01*
C4	2.56	1	0.11	17.18	8	0.03*
C5	0.46	1	0.50	9.21	8	0.33
C6	2.49	1	0.11	10.66	8	0.22
C7	1.81	1	0.18	10.76	8	0.22
C8	0.13	1	0.72	7.09	8	0.53
C9	0.31	1	0.58	13.87	8	0.09
C10	0.27	1	0.61	15.88	8	0.04*

* An asterisk indicates a significant P value below 0.05.

2.4.3.2 8M descriptive data

Mean component scores ranged from 6.92 for Separation to 84.03 for Energy; Table 2.13). All components had a minimum score of 0 except for C1 at 0.29 and Component 7 at 3.50. All components had a maximum score of 100 except for C2 and C3. Scores at 8M suggest that as a population, puppies on average scored highest for Energy. After Energy, the highest mean scores were reported for C6 at 40.50 (Attachment) and C1 (Impulsivity) at 36.31. See Appendix 8.4.2 for component distributions.

Table 2.13 8M r-PWQ2 component scores. Minimum, maximum, mean ($\pm$ standard error) for 10 behavioural components.

C	8M r-PWQ2	Min	Max	Mean	$\pm$ S.E.
C1	Impulsivity	0.29	100.00	36.31	0.44
C2	Anxiety	0.00	96.75	10.56	0.29
C3	Separation	0.00	95.25	6.92	0.25
C4	Distraction	0.00	100.00	30.75	0.37
C5	Anticipatory Excitement	0.00	100.00	9.82	0.37
C6	Attachment	0.00	100.00	40.50	0.49
C7	Energy	3.50	100.00	84.03	0.38

C8	Focus	0.00	100.00	22.34	0.37
C9	Body Sensitivity	0.00	100.00	11.19	0.36
C10	Attention Seeking	0.00	100.00	11.78	0.36

* A lower score is more desirable and '0' indicates the dog never performs the behaviour, except for C7 where no score is desirable.

Four components (Impulsivity, Anxiety, Attachment and Energy) were significantly impacted by sex and four components (Anxiety, Attachment, Focus and Attention Seeking) were significantly impacted by breed (Table 2.14). For a breakdown of mean component score by breed and sex see Table 8.11 in Appendix 8.5.2. Females had more desirable r-PWQ2 scores in three components, compared to males having more desirable scores in seven components. Favourable r-PWQ2 component scores varied by breed, with the GR (n=249) and the SP x Lab and Lab x SP group (n=29) having the highest number of desirable scores when compared to other breeds.

Table 2.14 Sex and breed effects for 8M behavioural components, reporting Wald Chi-square, degrees of freedom (df) and P-value.

C	Sex effect			Breed effect		
	Wald Chi-Square	df	P value	Wald Chi-Square	df	P-value
C1	4.21	1	0.04*	13.03	9	0.16
C2	5.91	1	0.02*	64.98	9	<0.001**
C3	3.83	1	0.05	6.89	9	0.65
C4	1.23	1	0.27	10.60	9	0.30
C5	0.26	1	0.61	14.59	9	0.10
C6	4.86	1	0.03*	22.02	9	<0.01*
C7	11.57	1	<0.001**	20.89	9	0.13
C8	0.16	1	0.69	19.44	9	0.02*
C9	1.72	1	0.19	15.95	9	0.07
C10	0.11	1	0.92	20.30	9	0.02*

* An asterisk indicates a significant P value below 0.05 and ** a P value below 0.001.

2.4.3.3 12M descriptive data

Mean component scores ranged from 4.74 for Separation to 84.01 for Energy; Table 2.15). All components had a minimum score of 0 and six

components had a maximum score of 100. The lowest maximum score 78.00 for C10 (Focus). Scores at 12M suggest that as a population, puppies on average scored highest for Energy. After Energy, the highest mean scores were reported for C3 at 29.19 (Impulsivity) and C4 (Focus) at 28.85. See Appendix 8.4.3 for component distributions.

Table 2.15 12M r-PWQ2 component scores. Minimum, maximum, mean ($\pm$ standard error) for 10 behavioural components.

C	12M r-PWQ2	Min	Max	Mean	$\pm$ S.E.
C1	Separation	0.00	100.00	4.74	0.24
C2	Anxiety	0.00	96.50	9.91	0.35
C3	Impulsivity	0.00	100.00	29.19	0.55
C4	Focus	0.00	97.17	28.85	0.44
C5	Attachment and Attention Seeking	0.00	100.00	28.09	0.46
C6	General Excitement	0.00	92.60	18.99	0.39
C7	Anticipatory Excitement	0.00	100.00	7.99	0.41
C8	Energy	0.00	100.00	84.01	0.47
C9	Body Sensitivity	0.00	100.00	11.05	0.45
C10	Human Focus	0.00	78.00	10.98	0.32

*A lower score is more desirable and '0' indicates the dog never performs the behaviour, except for C8 where no score is desirable.

No components were significantly impacted by sex and three components (Anxiety, Energy and Body Sensitivity) were significantly impacted by breed (Table 2.16). For a breakdown of mean component score by breed and sex see Table 8.13 in Appendix 8.5.3. Females had more desirable r-PWQ2 scores in six components, compared to males having more desirable scores in four components. Favourable r-PWQ2 component scores varied by breed, with the SP x Lab and Lab x SP group (n=21) having the highest number of desirable scores when compared to other breeds.

Table 2.16 Sex and breed effects for 12M behavioural components, reporting Wald Chi-square, degrees of freedom (df) and P-value.

C	Sex effect			Breed effect		
	Wald Chi-Square	df	P value	Wald Chi-Square	df	P-value
C1	1.04	1	0.31	6.13	9	0.73
C2	0.78	1	0.38	43.6	9	<0.001**
C3	0.30	1	0.58	10.60	9	0.30
C4	0.11	1	0.74	16.09	9	0.07
C5	1.80	1	0.18	6.63	9	0.68
C6	0.81	1	0.37	10.19	9	0.34
C7	1.08	1	0.30	7.86	9	0.55
C8	0.07	1	0.79	25.50	9	<0.01*
C9	0.00	1	1.00	25.79	9	<0.01*
C10	0.09	1	0.77	14.66	9	0.10

* An asterisk indicates a significant P value below 0.05 and ** a P value below 0.001.

2.4.3.4 Summary of scores

There were some differences in the breeds represented between 5M, 8M and 12M samples but no difference in the numbers of males and females represented. Many components had the full range of data reported with a minimum of 0 and maximum of 100. For all questionnaire time points, Separation had the lowest overall mean scores (8.8 at 5M, 6.9 at 8M and 4.7 at 12M). Energy had the highest mean score compared to all other components (84.9 at 5M, 84.0 at 8M and 84.01 at 12M). After Energy, the highest average scores were reported for C1 (Distraction and Human Focus) and C6 (Impulsivity) at 5M, for C6 (Attachment) and C1 (Impulsivity) at 8M and for C3 (Impulsivity) and C4 (Focus) at 12M.

In general, males tended to have more desirable r-PWQ2 scores at both 5M and 8M, but at 12M females had more components with desirable scores. Sex significantly impacted one component at 5M, four components at 8M and no components at 12M. Four components at both 5M and 8M, and three components at 12M were significantly impacted by breed. GR, SP x Lab grouped with Lab x SP and the Lab x GR* grouped with Lab x Lab* had the most desirable scores in general for components at 5M, 8M and 12M.

2.4.4 Comparing the r-PWQ and r-PWQ2

2.4.4.1 *Correlation between r-PWQ and r-PWQ2 components*

The r-PWQ and r-PWQ2 components were compared, and significant correlations were reported for all components at all time points. At 5M correlations coefficients ranged from 1.000 to -0.582 (Table 2.17). All correlation coefficients had a P value <0.01. The highest correlations were found for the Anxiety, Energy and Body Sensitivity components. The lowest correlation was reported for r-PWQ2 Focus and r-PWQ Trainability.

Table 2.17 r-PWQ and r-PWQ2 component correlation at 5M

C	5M r-PWQ2	Correlating r-PWQ component	Correlation coefficient
C1	Distraction and Human Focus	Distractibility	0.913**
		Trainability	-0.687**
C2	Separation	Separation-Related Behaviour	0.996**
C3	Anxiety	General Anxiety	1.000**
C4	General Excitement	Excitability	0.905**
C5	Attachment and Attention Seeking	Attachment and attention seeking	0.946**
C6	Impulsivity	Animal Chase	0.774**
		Distractibility	0.899**
C7	Energy	Energy Level	1.000**
C8	Anticipatory Excitement	Excitability	0.617**
C9	Body Sensitivity	Body Sensitivity	1.000**
C10	Focus	Trainability	-0.582**

**Correlation is significant at the 0.01 level (2-tailed).

At 8M correlations coefficients ranged from 1.000 to 0.597 (Table 2.18). All correlation coefficients had a P value <0.01. The highest correlations were found for the Anxiety, Energy and Body Sensitivity components. The lowest correlation was reported for r-PWQ2 Attention Seeking and r-PWQ Attachment and attention seeking.

Table 2.18 r-PWQ and r-PWQ2 component correlation at 8M

C	8M r-PWQ2	Correlating r-PWQ component	Correlation coefficient
C1	Impulsivity	Excitability	0.636**

		Animal Chase	0.737**
		Distractibility	0.884**
C2	Anxiety	General Anxiety	1.000**
C3	Separation	Separation-Related Behaviour	0.970**
C4	Distraction	Trainability	-0.826**
C5	Anticipatory Excitement	Excitability	0.650**
C6	Attachment	Attachment and attention seeking	0.877**
C7	Energy	Energy Level	1.000**
C8	Focus	Trainability	-0.670**
C9	Body Sensitivity	Body Sensitivity	1.000**
C10	Attention Seeking	Attachment and attention seeking	0.597**

**Correlation is significant at the 0.01 level (2-tailed).

At 12M correlations coefficients ranged from 1.000 to -0.553 (Table 2.19). All correlation coefficients had a P value <0.01. The highest correlations were found for the Anxiety, Energy and Body Sensitivity components. The lowest correlation was reported for r-PWQ2 Human Focus and r-PWQ Trainability.

Table 2.19 r-PWQ and r-PWQ2 component correlation at 12M

C	12M r-PWQ2	Correlating r-PWQ component	Correlation coefficient
C1	Separation	Separation-Related Behaviour	0.991**
C2	Anxiety	General Anxiety	1.000**
C3	Impulsivity	Animal Chase	0.800**
		Distractibility	0.904**
C4	Focus	Distractibility	0.813**
		Trainability	-0.773**
C5	Attachment and Attention Seeking	Attachment and attention seeking	0.940**
C6	General Excitement	Excitability	0.949**
C7	Anticipatory Excitement	Excitability	0.642**
C8	Energy	Energy Level	1.000**
C9	Body Sensitivity	Body Sensitivity	1.000**
C10	Human Focus	Trainability	-0.553**

**Correlation is significant at the 0.01 level (2-tailed).

2.4.4.2 Internal reliability of r-PWQ and r-PWQ2

The internal reliability of r-PWQ2 and the corresponding r-PWQ components were compared at 5M (Table 2.20), 8M (Table 2.21) and 12M (Table 2.22). At 5M, the internal reliability of r-PWQ2 components were generally higher than the corresponding r-PWQ components, but differences were small. The greatest difference was seen for Distraction and Human Focus r-PWQ2 (0.83) compared with r-PWQ Trainability (0.61). Internal reliability was higher for r-PWQ Trainability (0.61) when compared to r-PWQ2 Focus (0.47).

Table 2.20 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 5M

5M r-PWQ2	Correlating r-PWQ component	5M r-PWQ2 Cronbach's α	5M r-PWQ Cronbach's α
Distraction and Human Focus	Distractibility	0.83	0.84
	Trainability		0.61
Separation	Separation-Related Behaviour	0.84	0.82
Anxiety	General Anxiety	0.85	0.85
General Excitement	Excitability	0.74	0.72
Attachment and Attention Seeking	Attachment and attention seeking	0.65	0.62
Impulsivity	Animal Chase	0.80	0.75
	Distractibility		0.84
Energy	Energy Level	0.85	0.85
Anticipatory Excitement	Excitability	0.78	0.72
Body Sensitivity	Body Sensitivity	0.78	0.78
Focus	Trainability	0.47	0.61

At 8M, the internal reliability of r-PWQ2 components were generally higher than the corresponding r-PWQ components (Table 2.21). The greatest differences were for r-PWQ2 Anticipatory Excitement (0.82) compared with r-PWQ Excitability (0.73) and for r-PWQ2 Distraction (0.71) compared with r-PWQ Trainability (0.62). Internal reliability was higher for r-PWQ Attachment and attention seeking (0.60) when compared to r-PWQ2 Attachment (0.58).

Table 2.21 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 8M

8M r-PWQ2	Correlating r-PWQ component	8M r-PWQ2 Cronbach's α	8M r-PWQ Cronbach's α
Impulsivity	Excitability	0.84	0.73
	Animal Chase		0.77
	Distractibility		0.82
Anxiety	General Anxiety	0.83	0.83
Separation	Separation-Related Behaviour	0.76	0.74
Distraction	Trainability	0.71	0.62
Anticipatory Excitement	Excitability	0.82	0.73
Attachment	Attachment and attention seeking	0.58	0.60
Energy	Energy Level	0.85	0.85
Focus	Trainability	0.51	0.62
Body Sensitivity	Body Sensitivity	0.76	0.76
Attention Seeking	Attachment and attention seeking	0.63	0.60

At 12M, the internal reliability of r-PWQ2 components were generally higher than the corresponding r-PWQ components (Table 2.22). The greatest differences were for r-PWQ2 Anticipatory Excitement (0.86) compared with r-PWQ Excitability (0.72) and for r-PWQ2 Focus (0.76) compared with r-PWQ Trainability (0.62). Internal reliability was higher for r-PWQ Distractibility (0.80) when compared to r-PWQ2 Focus (0.76) and for r-PWQ Trainability (0.62) when compared to r-PWQ2 Human Focus (0.34).

Table 2.22 Internal reliability for corresponding r-PWQ and r-PWQ2 components at 12M

12M r-PWQ2	Correlating r-PWQ component	12M r-PWQ2 Cronbach's α	12M r-PWQ Cronbach's α
Separation	Separation-Related Behaviour	0.80	0.79
Anxiety	General Anxiety	0.83	0.83
Impulsivity	Animal Chase	0.82	0.79
	Distractibility		0.80
Focus	Distractibility	0.76	0.80
	Trainability		0.62

Attachment and Attention Seeking	Attachment and attention seeking	0.64	0.60
General Excitement	Excitability	0.66	0.72
Anticipatory Excitement	Excitability	0.86	0.72
Energy	Energy Level	0.81	0.81
Body Sensitivity	Body Sensitivity	0.77	0.77
Human Focus	Trainability	0.34	0.62

2.4.4.3 r-PWQ and r-PWQ2 conclusions

Following the correlation analysis between the r-PWQ and the r-PWQ2, significant correlations provided confidence in the data structure and content validity of the r-PWQ2. The internal consistency between r-PWQ and r-PWQ2 components was also comparable, with a trend for a higher internal reliability for the r-PWQ2 components when compared to the r-PWQ. As such the r-PWQ2 data was utilised in the remainder of the thesis.

2.4.5 Additional factors

2.4.5.1 Puppy raiser information

The 4405 questionnaires were completed by 1921 puppy raisers. One puppy raiser for one 8M questionnaire was unknown. Puppy raisers could have raised more than one puppy during data collection so individual puppy raisers completed varying numbers of questionnaires depending on the number of puppies raised. Five hundred and seventy-eight (30.1%) puppy raisers completed one questionnaire, 724 (37.7%) completed two questionnaires, 290 (15.1%) puppy raisers completed three questionnaires, 195 (10.2%) puppy raisers completed four questionnaires, and the remaining 134 (7.0%) puppy raisers completed over five questionnaires in the sample. Three puppy raisers completed eight questionnaires.

Puppy raiser experience ranged from this being the first puppy to 53 puppies being raised by a volunteer (Table 2.23). There was no difference in the experience of puppy raisers between the three time points ($\chi^2=5.81$, D.F.=6, $P=0.44$). At 5M, puppy raiser experience ranged from this being the first puppy ($n=124$) to raising 53 puppies ($n=1$). The average number of puppies raised by puppy raisers was 6.2. At 8M, puppy raiser experience ranged from this being the first puppy ($n=470$) to raising 53 puppies ($n=1$). The average number of puppies raised by puppy raisers was 5.4. At 12M, puppy raiser experience ranged from this being the first puppy raised ($n=289$) to raising 53 puppies ($n=1$). The average number of puppies raised was 5.8.

Table 2.23 Puppy raiser experience for data collected at five (5M), eight (8M) and 12 (12M) of age.

Puppy raiser experience	N (%) 5M puppy raisers	N (%) 8M puppy raisers	N (%) 12M puppy raisers
1-4 puppies	323 (58.6%)	1359 (58.7%)	881 (57.3%)
5-9 puppies	129 (23.4%)	523 (22.6%)	351 (22.8%)
10-19 puppies	65 (11.8%)	328 (14.2%)	226 (14.7%)
20 + puppies	34 (6.2%)	103 (4.4%)	78 (5.1%)
Boarder or unknown	0 (0.0%)	4 (0.2%)	1 (0.1%)

The 2892 dogs were supervised by 65 PDAs. The number of puppies supervised by each PDA ranged from one to 160 (mean 67.8 puppies ± 41.6).

2.4.5.2 *Puppy raiser home environment*

Information on additional factors was collected from puppy raisers while the r-PWQ2 was completed (Table 2.24). The differences between data collected at each of the three time points were examined. The dogs with missing information were not included in analyses and the data available varied between each factor (Table 2.25).

Table 2.24 Additional factor information for r-PWQ2 questionnaires. Number and percentage of factors for data collected at five (5M), eight (8M) and 12 (12M) of age.

		5M	8M	12M
Number of adults	One adult	80 (14.8%)	280 (14.2%)	222 (14.7%)
	Two adults	356 (65.9%)	1311 (66.3%)	975 (64.5%)
	Three adults	76 (14.1%)	269 (13.6%)	214 (14.2%)
	Four adults	21 (3.9%)	82 (4.1%)	76 (5.0%)
	Five or more	7 (1.3%)	34 (1.7%)	25 (1.7%)
Presence of children	Yes	108 (27.6%)	410 (29.6%)	290 (27.4%)
	No	284 (72.4%)	974 (70.4%)	767 (72.6%)
Presence of dogs	Yes	197 (39.7%)	744 (38.5%)	559 (38.0%)
	No	299 (60.3%)	1186 (61.5%)	921 (62.0%)
Presence of animals	Yes	274 (50.8%)	1026 (53.1%)	778 (52.5%)
	No	265 (49.2%)	905 (46.9%)	703 (47.5%)
Puppy raiser home location	Rural	175 (31.9%)	617 (31.0%)	466 (30.6%)
	Suburban	291 (53.1%)	1051 (52.7%)	815 (53.4%)
	Urban	82 (15.0%)	325 (16.3%)	244 (16.0%)
Puppy class	Yes	450 (82.3%)	1542 (90.4%)	689 (90.2%)
	No	97 (17.7%)	164 (9.6%)	75 (9.8%)
Workplace	Yes	62 (11.3%)	211 (12.3%)	108 (14.2%)
	No	485 (88.7%)	1500 (87.7%)	651 (85.8%)
Year of Birth	2016	0 (0.0%)	304 (13.1%)	33 (2.1%)
	2017	0 (0.0%)	886 (38.2%)	730 (47.5%)
	2018	0 (0.0%)	674 (29.1%)	652 (42.4%)
	2019	551 (100.0%)	453 (19.6%)	122 (7.9%)
Neutered	Yes	NA	403 (18.6%)	991 (68.2%)
	No	NA	1765 (81.4%)	462 (31.8%)

Table 2.25 Additional factor missing information for r-PWQ2 questionnaires. Number and percentage of missing information for data collected at five (5M), eight (8M) and 12 (12M).

Factor	5M	8M	12M
Adults	11 (2.0%)	341 (14.7%)	25 (1.6%)
Children	159 (28.9%)	933 (40.3%)	480 (31.2%)
Dogs	55 (10.0%)	387 (16.7%)	57 (3.7%)
Animals	12 (2.2%)	386 (16.7%)	56 (3.6%)
Home location	3 (0.5%)	324 (14.0%)	12 (0.8%)
Puppy Class	4 (0.7%)	611 (26.4%)	773 (50.3%)
Workplace	4 (0.7%)	606 (26.2%)	778 (50.6%)
Neutered	N/A	149 (6.4%)	84 (5.5%)

There was no difference in the number of adults in the home between the three time points ($\chi^2=2.62$, D.F.=6, $P=0.85$). Households were most likely

to have two adults present and were least likely to have four or more adults. There was no difference in the number of children in the home between the three time points ($\chi^2=1.62$, D.F.=2, $P=0.45$). Under 20% of puppy raising homes were reported to have children living in the home, however the percentage of missing data was quite high for all time points (28.9%, 40.3% and 31.2%). There was no difference in the number of homes where another dog was present between the three time points ($\chi^2=0.63$, D.F.=2, $P=0.73$). More than half of puppy raising homes did not have a dog at any of the three times points. There was no difference in the number of homes that had other animals (not including dogs) present between the three time points ($\chi^2=0.90$, D.F.=2, $P=0.64$). There was no difference in the location of the puppy raisers home between the three time points ($\chi^2=0.81$, D.F.=4, $P=0.94$). Puppy raisers typically reported living in suburban locations (53.1% at 5M, 52.7% at 8M and 53.4% at 12M). There was a difference in the number of puppy raisers that reported attending puppy classes between the three time points ($\chi^2=29.40$, D.F.=2, $P<0.001$). There was a difference between the 5M and 8M data sets ($\chi^2=26.66$, D.F.=1, $P<0.001$) and between the 5M and 12M data sets ($\chi^2=17.53$, D.F.=1, $P<0.001$). There was no difference between 8M and 12M data sets ($\chi^2=0.03$, D.F.=1, $P=0.87$). There was no difference in the number of puppy raisers that reported taking their puppy into the workplace between the three time points ($\chi^2=2.73$, D.F.=2, $P=0.26$).

At 8M, data on neutering age were available for 2168 puppies (93.6% of the eight-month sample), with 403 dogs being neutered before the questionnaire expiry (8.5 months of age) and 1765 after. Average neutering age for those neutered before questionnaire expiry at 8M was 8.14 months (± 0.28 months) and for those after the questionnaire expiry was 12.04 months (± 2.81 months). At 12M, data on neutering age were available for 1453 puppies (94.5% of the 12M sample), with 991 dogs being neutered before the questionnaire expiry (12.5 months of age) and 462 after. Average neutering age for those neutered before questionnaire expiry at 12M was 9.56 months (± 1.47 months) and for those after the questionnaire expiry was 14.50 months (± 2.16 months).

2.4.5.3 5M factor influence

For each r-PWQ2 component, all factors were tested for the effect on component scores (Table 2.26). Litter impacted the most components (80% of components) for all the factors studied. All components except Distraction and Human Focus and Impulsivity were impacted by litter. The PDA supervising the puppy, the puppy raiser experience, and the presence of children in the home impacted 60% of the components. The presence of dogs in the home impacted Separation, Anxiety and Energy components and the location of the puppy raisers home impacted Attachment and Attention Seeking and Anticipatory Excitement. Only one component was impacted by attendance at a workplace (Anticipatory Excitement). No components were significantly impacted by the number of adults in the home, presence of animals in home or attendance at puppy classes. Body Sensitivity was impacted by the least number of factors (n=1) and Anxiety and Anticipatory Excitement were both impacted by five factors, the highest number of factors to significantly impact a component score.

Table 2.26 Significant factors for r-PWQ2 components at 5M.

C	5M r-PWQ2	Significant factors (P<0.05)
C1	Distraction and Human Focus	PDA, PR experience ⁽⁺⁾ , Children ⁽⁻⁾ .
C2	Separation	Litter, PR experience ⁽⁺⁾ , Dogs ⁽⁺⁾ .
C3	Anxiety	Litter, PDA, PR experience ⁽⁻⁾ , Children ⁽⁻⁾ , Dogs ⁽⁺⁾ .
C4	General Excitement	Litter, PDA, PR experience ⁽⁺⁾ , Children ⁽⁻⁾ .
C5	Attachment and Attention Seeking	Litter, PDA, PR experience ⁽⁺⁾ , Location (suburban).
C6	Impulsivity	PR experience ⁽⁺⁾ , Children ⁽⁻⁾ .
C7	Energy	Litter, PDA, Dogs ⁽⁻⁾ .
C8	Anticipatory Excitement	Litter, PDA, Children ⁽⁻⁾ , Location (suburban), Workplace ⁽⁺⁾
C9	Body Sensitivity	Litter
C10	Focus	Litter, Children ⁽⁻⁾ .

For PR experience ⁽⁺⁾ indicates a trend for more favourable behaviour with more experienced raisers. For the presence of children/ dogs, a ⁽⁺⁾ indicates more favourable behaviour and ⁽⁻⁾ less favourable behaviour when dogs/ children were present. The type of location with more favourable behaviour is indicated. Attendance at the workplace displaying more favourable behaviours is indicated

as ⁽⁺⁾. Due to the number of different PDAs and litters, the direction of favourable behaviour is not included.

Puppy raiser experience was shown in general to have more favourable r-PWQ2 scores linked with more puppies raised. For the six components impacted by PR experience (C1, C2, C3, C4, C5 and C6), five of the six components had the most favourable score when 20 or more puppies had been raised by the PR. The presence of children was shown to impact six components (C1, C3, C4, C6, C8 and C10), with puppies having more desirable scores when no children were living in the home. The presence of dogs influenced three components (C2, C3 and C7) with more desirable scores seen when no other dogs were in the home. Puppy raiser home location influenced C5 and C8, with more desirable r-PWQ2 scores seen when puppy raisers lived in a suburban location. Visiting the workplace has a positive influence on C8, with scores more favourable for puppies that had visited a workplace. See Appendix 8.5.1 for detailed r-PWQ2 scores for each factor.

2.4.5.4 8M factor influence

For each r-PWQ2 component, all factors were tested for the effect on component scores (Table 2.27). Litter impacted all components at 8M. PR experience impacted 80% of the components (all components except for Energy and Attention Seeking) and PDA impacted seven (70%) components. The presence of children and the location of the puppy raisers home impacted five components. No components were significantly impacted by visiting the workplace. The presence of dogs in the home, whether the puppy had been neutered, the presence of animals in the home or attendance at puppy classes were each significant for one component. The number of adults in the home impacted two components and all other factors impacted three or more components. Separation was impacted by the least number of factors (n=2) and Attachment impacted by the most factors (n=8).

Table 2.27 Significant factors for r-PWQ2 components at 8M.

C	8M r-PWQ2	Significant factors (P<0.05)
C1	Impulsivity	YOB ⁽²⁰¹⁷⁾ , Litter, PDA, PR experience ⁽⁺⁾ , Location ^(rural) , Neuter status ^(no)
C2	Anxiety	Litter, PR experience ⁽⁺⁾ , Adults ⁽⁺⁾ , Children ⁽⁻⁾ .
C3	Separation	Litter, PR experience ⁽⁺⁾ ,
C4	Distraction	YOB ⁽²⁰¹⁶⁾ , Litter, PDA, PR experience ⁽⁺⁾ , Location ^(suburban) .
C5	Anticipatory Excitement	Litter, PDA, PR experience ⁽⁺⁾ , Children ⁽⁻⁾ , Neuter status ^(yes)
C6	Attachment	YOB ⁽²⁰¹⁶⁾ , Litter, PDA, PR experience ⁽⁺⁾ , Children ⁽⁻⁾ , Animals ⁽⁺⁾ , Location ^(suburban) , Neuter status ^(no)
C7	Energy	Litter, PDA, Children ⁽⁺⁾ .
C8	Focus	Litter, PDA, PR experience ⁽⁻⁾ , Location ^(rural) .
C9	Body Sensitivity	YOB ⁽²⁰¹⁶⁾ , Litter, PR experience ⁽⁻⁾ , Puppy class ⁽⁻⁾ .
C10	Attention Seeking	Litter, PDA, Adults ⁽⁺⁾ , Children ⁽⁻⁾ , Dogs ⁽⁻⁾ , Location ^(suburban) .

For YOB, the year with more favourable scores is indicated e.g. ⁽²⁰¹⁷⁾. For PR experience ⁽⁺⁾ indicates a trend for more favourable behaviour with more experienced raisers. For the presence of children/dogs/other animals, a ⁽⁺⁾ indicates more favourable behaviour and ⁽⁻⁾ less favourable behaviour when dogs/ children/ animals were present. The type of location with more favourable behaviour is indicated. Attendance at puppy class resulting in more favourable behaviours is indicated as ⁽⁺⁾. For neutering ^(yes) indicates more favourable behaviour for neutering by questionnaire expiry and ^(no) for less favourable behaviour. Due to the number of different PDAs and litters, the direction of favourable behaviour is not included.

Puppy raiser experience was shown in general to have more favourable r-PWQ2 scores linked with more puppies raised. For the eight components impacted by PR experience (C1, C2, C3, C4, C5, C6, C8 and C9), five of the components had the most favourable score when 20 or more puppies had been raised by the PR. C8 and C9 showed more favourable scores for the less puppies raised. The presence of children was shown to impact five components (C2, C5, C6, C7 and C10), with puppies having more desirable scores when no children were living in the home for four of the five components. Location influenced five components (C1, C4, C6, C8 and C10) with a mixture of rural and suburban locations having favourable scores when compared to urban locations. Dogs in the home had a negative influence on C10 and attendance at a puppy class had a negative influence on C9. The presence of living with other animals positively influenced C6 scores. The number of adults in the home

influenced C2 and C10, with the presence of five or more adults positively influencing r-PWQ2 scores. Year of birth influenced four components (C1, C4, C6 and C9), with more favourable scores seen for dogs born in 2016 for three of the four components. Neuter status influenced three components (C1, C5, C6), with two of the three components having more favourable scores when not neutered. See Appendix 8.5.2 for detailed r-PWQ2 scores for each factor.

2.4.5.5 12M factor influence

For each r-PWQ2 component, all factors were tested for the effect on component scores (Table 2.28). Litter impacted all components at 12M. All components except for General Excitement, Energy and Body Sensitivity were impacted by PDA. The number of adults in the home and the whether the puppy had been neutered impacted 60% of components, whilst the experience of the puppy raiser impacted half of the components (50%). No components were significantly impacted by the location of the puppy raisers home or visiting the workplace. One component was impacted by attendance at puppy classes, the presence of dogs and the year of birth impacted two components, and all other factors impacted three or more components. C9 (Body Sensitivity) was impacted by the least number of factors (n=1) and C5 (Attachment and Attention Seeking) was significantly impacted by the greatest number of factors (n=10).

Table 2.28 Significant factors for r-PWQ2 components at 12M.

C	12M r-PWQ2	Significant factors (P<0.05)
C1	Separation	Litter, PDA, Adults ⁽⁺⁾ , Puppy class ⁽⁺⁾ .
C2	Anxiety	Litter, PDA, Adults ⁽⁻⁾ , Children ⁽⁻⁾ , Animals ⁽⁻⁾ , Neuter status ^(yes)
C3	Impulsivity	Litter, PDA, PR experience ⁽⁺⁾ , Children ⁽⁻⁾ , Neuter status ^(yes)
C4	Focus	Litter, PDA, PR experience ⁽⁺⁾ , Neuter status ^(yes)
C5	Attachment and Attention Seeking	YOB ⁽²⁰¹⁶⁾ , Litter, PDA, PR experience ⁽⁺⁾ , Adults ⁽⁻⁾ , Children ⁽⁻⁾ , Dogs ⁽⁻⁾ , Animals ⁽⁻⁾ , Neuter status ^(yes)
C6	General Excitement	Litter, PR experience ⁽⁺⁾ , Adults ⁽⁺⁾ , Neuter status ^(no)

C7	Anticipatory Excitement	Litter, PDA, Dogs ⁽⁻⁾ , Animals ⁽⁻⁾ , Neuter status ^(yes)
C8	Energy	Litter, PR experience ⁽⁺⁾ , Adults ⁽⁺⁾ ., Children ⁽⁺⁾ .
C9	Body Sensitivity	Litter.
C10	Human Focus	YOB ⁽²⁰¹⁶⁾ , Litter, PDA, Adults ⁽⁻⁾ .

For YOB, the year with more favourable scores is indicated e.g. ⁽²⁰¹⁷⁾. For PR experience ⁽⁺⁾ indicates a trend for more favourable behaviour with more experienced raisers. For the presence of children/dogs/other animals, a ⁽⁺⁾ indicates more favourable behaviour and ⁽⁻⁾ less favourable behaviour when dogs/ children/ animals were present. Attendance at puppy class resulting in more favourable behaviours is indicated as ⁽⁺⁾. For neutering ^(yes) indicates more favourable behaviour for neutering by questionnaire expiry and ^(no) for less favourable behaviour. Due to the number of different PDAs and litters, the direction of favourable behaviour is not included.

Puppy raiser experience was shown in general to have more favourable r-PWQ2 scores linked with more puppies raised. For the five components impacted by PR experience (C3, C4, C5, C6 and C8), more favourable scores were seen when more puppies had been raised by the PR. The number of adults in the home influenced six components (C1, C2, C5, C6, C8 and C10), with three components showing more desirable behaviours in the presence of 4 or more adults in the home, and the other three having more desirable behaviours when two adults were present in the home. Presence of children in the home impacted four components (C2, C3, C5 and C8), with three of these four components showing more desirable scores when no children were living in the home. Presence of another dog influenced C5 and C7 with more favourable scores for puppies when there were no other dogs in the home. Living with other animals influenced three components (C2, C5 and C7) with more desirable score when no other animals were present. Year of birth influenced C5 and C10 with more desirable scores seen for puppies born in 2016. Attendance at puppy class resulted in more favourable C1 scores. Neuter status influenced six components (C2, C3, C4, C5, C6 and C7), with five of the six components showing more favourable scores when neutered. See Appendix 8.5.3 for detailed r-PWQ2 scores for each factor.

2.4.5.6 Summary of additional factors

Most puppy raisers completed either one (30.1%) or two (37.7%) questionnaires in the whole data set and puppy raiser experience was similar between the questionnaire time points. There was a trend for puppy raiser experience to positively influence r-PWQ2 component scores when more puppies had been raised. Litter influenced the greatest number of components (n=8 at 5M, n=10 at 8M, n=10 at 12M) followed by the PDA influencing six components at 5M and seven components at 8M and 12M. The presence of children also impacted around half of components at each time point (n=6 at 5M, n=5 at 8M, n=4 at 12M), with a trend for scores to be more favourable when no children were in the home at all three time points. The presence of other dogs tended to have a positive influence on scores at 5M and a less favourable influence on scores at 8M and 12M. The same patterns were found for the presence of other animals in the home. Puppies raised in rural and suburban locations tended to present more favourable scores, this pattern was found at all three time points. Visiting a workplace (n=1 at 5M, n=0 at 8M, n=0 at 12M), and attendance at puppy classes (n=0 at 5M, n=1 at 8M, n=1 at 12M), were the factors with the least impact on component scores. Visiting a workplace and attendance at a puppy class provided a trend for more favourable scores at 5M in comparison to a trend for less desirable scores at 8M and 12M. More adults in the home influenced scores for younger puppies at 5M and 8M, whereas the presence of two adults demonstrated more desirable scores at 12M. Puppies born in 2016 tended to show more favourable scores when compared to puppies born in other years for the 8M and 12M samples. Whether neutered by the time of questionnaire expiry resulted in more desirable scores at 12M, but puppies that had not been neutered showed a trend for more desirable scores at 8M. In summary, different factors affected the scores at 5M, 8M and 12M, but similarities were noted (Table 2.29). The main factors affected component scores at all time points were litter, PDA, puppy raiser experience and the presence of children.

Table 2.29 Summary of factor significance for 5M, 8M and 12M components.

Factor reviewed	Number significant at 5M	Number significant at 8M	Number significant at 12M
Litter	8	10	10
Puppy Development Advisor	6	7	7
Puppy Raiser Experience	6	8	5
Number of adults	0	2	6
Live with children	6	5	4
Live with dogs	3	1	2
Live with animals	0	1	3
House location	2	5	0
Attended puppy class	0	1	1
Workplace	1	0	0
Year of birth	N/A	4	2
Age at Neutering	N/A	3	6

2.5 Discussion

The main objectives of this chapter were to re-examine the distribution and use of a behavioural questionnaire for a cohort of guide dog puppies. Previous work has shown the benefit of collecting this information from the volunteers caring for puppies during their first year of life (Harvey, 2014; Hunt *et al.*, 2020). Results from the present data suggest that the structure and grouping of behavioural components differ slightly for some components when compared to the original groupings reported by Harvey (2014). The main differences were the formation of 10 components compared to the nine for the r-PWQ, and the removal of some questions in the r-PWQ2 (two questions at 5M, six at 8M and three at 12M). The r-PWQ and r-PWQ2 components are not directly comparable due to slight differences in the questionnaire structure, relevance to the specific population and questionnaire purpose.

The original study plan was to collect data for the same puppies at all three questionnaire time points. Unfortunately, this was not possible due to the Covid-19 pandemic, that occurred soon after the study commenced. Due to the potential impact of the pandemic on dog behaviour (Gray,

2020), including impacts of social isolation and reduced environmental exposure on dog behaviour, it was not possible to collect reliable and valid data during this time. Therefore, a decision was made to use a mixture of retrospective (data collected in the past) and prospective data (planned data collection in the future), so the same cohort of dogs did not feature throughout all the data. The study could have been improved if puppies had data collected at all three time points. There were, however, around 50% of puppies that had more than one questionnaire completed within the data set, so multiple questionnaires for individuals did feature in the study.

The original data set for PWQ development was considerably smaller than the one used in the present study. Harvey (2014) collected data for 263 puppies at 5M, 214 puppies at 8M and 226 puppies at 12M. Furthermore, puppies included in the 2014 study were born earlier (in 2012) whereas puppies in the present study were born after 2015. There may have been a shift in the presence of certain behaviours over the years due to selective breeding. The two studies were not directly comparable due to the different population sizes, and the 2014 study contained more questions for the PCA as the study was centred around designing a questionnaire for use in a guide dog population. It may be that questions not deemed suitable for inclusion within the original study could have been appropriate for this population, a factor that is worthy to note when discussing the findings and for future use of the r-PWQ2. Since the Harvey (2014) study, GDUK have invested heavily in standardising training, ensuring all practices are positive reinforcement based and utilising different language and terminology when describing dog behaviour. It is also possible that the puppy raisers involved in the present study had different levels of experience or education relating to dog behaviour, which may impact how behaviour is measured. Whether the same puppy raisers featured in both studies has not been investigated.

When data were collected for the original study (Harvey, 2014), questions were asked in a randomised order. Following implementation at GDUK, the questions were asked in the order in that they appear within groupings. Therefore, there is the potential for a previous question to influence the next question. This was evidenced by the findings for the “is self-controlled and calm” question when asked twice. The first time the question was asked, it was surrounded by negative weighted questions, and the second time by positive weighted questions. It is likely the placement of these question could have affected the scores chosen. The moderate rather than high correlations observed between the two scores for this question ($r_s=0.60$ at 5M, $r_s=0.62$ at 8M and $r_s=0.63$ at 12M) suggests that the placement of the questions in a questionnaire is important to consider during development. Question order has been noted to influence responses, and thought should be given to the order in which they are asked (Krosnick, 2018). Some questionnaires randomise the order of questions; however, it has been suggested that items on related topics should be grouped together to support the flow of the questionnaire and so there is a logical order (Malhotra, 2006; Krosnick, 2018). GDUK could consider randomising the order of some questions, but it is unknown what impact this may have on future responses and thus may impact the comparison between different cohorts of dogs.

The questionnaires were analysed using a PCA utilising a varimax rotation. The varimax rotation was developed by Kaiser (1958) and is the most commonly used rotation within PCAs, where the results for each component have a small number of large loadings and a large number of zero (or small) loadings (Abdi and Williams, 2010). Jolliffe and Cadima (2016) state that it is common to use 70% total variance for the data set as a cut off point for components. In the present study, the variation explained by the extracted components accounted for approximately 65% of the variation in all data sets (64.7% at 5M, 64.1% at 8M and 61.5% at 12M), which is just below the reported suggested value. Harvey (2014) also utilised a varimax rotation and the PCA components accounted for

between 65 and 71% of the variance reported in the data sets, slightly higher than the findings in the present study. Schreiber (2021) noted that when having a large sample size (over 300), lower figures are acceptable, suggesting appropriate use in the present study.

For each PCA undertaken (5M, 8M and 12M) 10 components were extracted. Harvey (2014) reported between 15 and 16 components, although only nine scales were reported in the final questionnaire. It is common for differences to be found in the grouping of questionnaires. For example, Nagasawa *et al.* (2016) reported 63 of 100 C-BARQ questionnaire items to be common between two countries (Japan and USA) for factor analysis, with 12 factors emerging in over 13,000 dogs, compared with 14 factors reported in the original C-BARQ. The study also reported several questions to have response rates below 85% and these questions were excluded from further analyses, as were questions with factor loadings below 0.4.

Another difference between the PCA utilised by this study and that by Harvey (2014) is the primary use of the questionnaire. Harvey (2014) designed the questionnaire to be a measure of personality in guide dog puppies, and to predict where puppies may not succeed as guide dogs. The purpose of the questionnaire in this present study was to review heritabilities (both pedigree-based and genomic-based). It was appropriate for Harvey to use the same questions and groupings for each of the three time points based on the purpose of the questionnaire. Harvey (2014) noted in the discussion how “the PCA results presented here showed inconsistent patterns, with many component structures changing at each age.” The current study also found there to be differences in the structures between the three time points, particularly more differences noted at 8M. It was appropriate to use different questionnaire groupings at each time point to ensure the data were constructed in the best way possible for future analyses reviewing heritability. As such a new questionnaire was used, the r-PWQ2. It is important to remember that the components are

not directly comparable for all three time points, and caution should be used when describing the r-PWQ2 components at 5M, 8M and 12M.

By having differences between the questionnaire groupings at the three times, the construct validity of the r-PWQ2 is weakened. Furthermore, the r-PWQ2 has not been validated against other behavioural measures which adds further limitations. Conversely, when using the r-PWQ groupings for this data set, results suggested that two traits displayed a lower than acceptable Cronbach's alpha below 0.7 (Tavakol and Dennick, 2011). Correlations between the r-PWQ and r-PWQ2 components were significant for all traits and all time points supporting validity in the data. Furthermore, in general, the internal reliability of the r-PWQ2 components were higher for this data set than the comparable r-PWQ components, strengthening the justification to use the r-PWQ2 in this thesis.

Taylor and Mills (2006) defined internal consistency as the "correlation between items within each factor, where factors are derivations of behavioural observations obtained by data reduction techniques". Once the components had been determined and extracted, the internal reliability of the components were reviewed using Cronbach's alpha.

Cronbach's alpha of greater than 0.70 is considered to indicate good internal reliability (Tavakol and Dennick, 2011), but scales with fewer than 10 items can be considered acceptable over 0.5 (Taber, 2018). The internal reliability of the original scales ranged from 0.34-0.85 at 5M, 0.24-0.87 at 8M and 0.31-0.89 at 12M (Harvey, 2014). The internal reliability of the r-PWQ2 components in this study ranged from 0.47-0.85 at 5M, 0.51-0.85 at 8M and 0.34-0.86 at 12M. The reliability of the scales in the present study was generally higher than the original study, which may be due to the larger sample used in analyses.

A review of the questions and terminology used in the r-PWQ could be explored. For example, questions for the Anxiety component contain broad terms such as 'disturbed', 'spooked' or being 'anxious or uneasy'. Whilst

these questions use less subjective terminology than previous data collected at GDUK, suggested improvements include providing a definition to ensure that interpretation between volunteers is as consistent as possible. Other behaviours such as vocalising (whining, barking), lip licking, body shaking, panting, a lowered body posture, freezing and displacement (e.g. sniffing, scratching, chewing) have all been associated with fear-related behaviours in dogs (Stellato *et al.*, 2017), yet these behaviours have not been referred to in the r-PWQ questions. Since implementation of the r-PWQ at GDUK, additional questions have been added to the questionnaire to ensure a wide range of behaviours are being captured during puppy raising. The additional questions are worded to be like r-PWQ questions and use the same visual analogue scale of 0-100. Future studies, outside of the remit of the present study, will review the components and structure of the questionnaire with the addition of the new questions.

Puppy raiser experience was shown to be a significant factor in many different component scores and could indicate the importance of puppy raiser learning and education. The general trend was for more desirable r-PWQ2 scores for puppy raisers that had raised more puppies. GDUK have recently reviewed the puppy raising programme and the importance of standardising information provided to puppy raisers. As such the Puppy Raising for Excellent Partnerships (PREP) programme was launched at the end of 2021. Before this, the puppy raising level of supervision and advice would vary by region and PDA. The puppy raisers' previous experience raising dogs prior to puppy raising for GDUK was unknown and is a factor that could have impacted component scores and could be included in future studies.

Future research could determine correlations between human and dog personality using the r-PWQ2. It has been suggested that as humans and dogs coexist together in many environments, there may some similarity in human and dog personality (Chopik and Weaver, 2019). Turcsán *et al.*

(2012) reported positive correlations between owners and their dogs in five personality dimensions (neuroticism, extraversion, conscientiousness, agreeableness, and openness). Similarly, a study of 555 pet owners found that more positive attitudes towards dogs and cats were reported when the behavioural styles of the pets complemented their owners (Zeigler-Hill and Highfill, 2010). These results might be expected as humans commonly choose dog breeds and personalities that would complement their lifestyle such as activity levels, companionship and intelligence (Chopik and Weaver, 2019).

Environment and owner demographics or characteristics have been shown to influence dog behaviour. Kubinyi *et al.* (2009) reviewed owner demographic characteristics and dog personality associations in over 14,000 owners using four personality traits in the dog (Calmness, Trainability, Dog sociability, Boldness). Calmness was shown to be higher for dogs owned by men, and when other dogs were in the house. In the present study, although puppy raiser sex was not considered, a variety of different associations were found between home environment factors and r-PWQ2 behaviour components. The factors with the greatest impact were found to be litter, PDA and puppy raiser experience. Some factors had a small impact on the component scores at any time point, such as visiting the workplace and attendance at puppy class. Whether the puppy had been neutered before questionnaire expiry window for 8M and 12M data was reviewed. However, there were limitations using this data and whether the puppy had or had not been neutered at the time of questionnaire completion should have been calculated. This factor should also have been included in the breed and sex influences model rather than a review of additional factors. Harvey *et al.* (2016) similarly reported the age of the puppy raiser, number of puppies previously raised, and the number of dogs and children present in the household to significantly impact PWQ behaviour scores, but a smaller number of factors were reviewed compared to the present study so direct comparisons cannot be made.

2.6 Conclusion

This chapter aimed to determine whether data collected through a validated behavioural questionnaire, the r-PWQ, had the same questionnaire and component structure for questionnaires collected at three time points (5M, 8M and 12M) when compared to the original design of the questionnaire. Results suggest some differences in the grouping of questions within the components extracted compared to the original development of the questionnaire by (Harvey, 2014) and differences in the factors effecting component scores between the time points.

The research question was answered, and the findings suggest that the use of the pre-existing questionnaire structure for the present study population is not suitable. Some questions were not relevant for inclusion in the components, and there were differences between the three time points noted in terms of which questions were placed within each component. Ten components were extracted at each time point; however, some were shown to be unreliable due to the low internal consistency of the questions when grouped together. Due to the large impact of the puppy raiser and their experience on the behaviour of puppies, future studies using behavioural questionnaires could consider associations between human personality and dog temperament.

Chapter 3. Heritability analysis of a behavioural questionnaire (the r-PWQ2) for guide dog puppies at five, eight and 12 months of age

3.1 Introduction

The goal of selective breeding is to improve the average genetic merit of a population, thereby increasing the likelihood of higher-performing animals in the next generation (Chen *et al.*, 2021a). By selecting and breeding animals that are behaviourally suited for a desired role, there is the potential to influence the behaviour of future animals (Goddard and Beilharz, 1982; Joshi, 2005; Grandin and Deesing, 2013). Behaviour is complex and often a result of interactions between multiple genes rather than one specific gene (Plomin, 1990; Evans *et al.*, 2015), therefore determining the influences of breeding on behaviour becomes challenging. Understanding genetic influences on behaviour can support maximising welfare by producing animals less likely to display problem behaviours. For animals used within production and working roles, understanding factors that can improve efficiencies and aid selection can save considerable resources and enhance impact (e.g., more assistance dogs available to support people with disabilities). However, breeding for traits will only be useful if the traits are measured accurately and demonstrate genetic variation within the population (Ruefenacht *et al.*, 2002). It is therefore essential that breeding programmes focus on good data keeping, collection of unbiased and non-subjective data, and that pedigrees and parentage are recorded correctly.

Estimating heritability remains a practical and essential tool for breeding programmes (Hradecká *et al.*, 2015). Heritability (h^2) indicates the degree to which a specific trait or health condition may change through breeding selection within a population (Falconer and Mackay, 1989). Heritability has been described as the ratio of genetic to phenotypic variance (Falconer and Mackay, 1989) and can be represented in simple terms as: $h^2 = \frac{V_g}{V_p}$

where, V_g is the genetic variance and V_p the phenotypic variance. The heritability therefore denotes the influence of genetics on that trait. High heritabilities suggest that variation observed is caused by genotype variation, with less environmental influence and more resemblance is seen between relatives (Visscher *et al.*, 2008). Heritability is reported as a proportion of 1, with 0 indicating the trait is 0% heritable, and 1 the trait is 100% heritable. Traits with heritabilities below 0.2 are considered to be low, between 0.2 and 0.4 are considered moderate and heritabilities greater than 0.4 are considered high (Bourdon, 2000). Heritability can be calculated by pedigree-based methods (utilising a relationship matrix for the population) or genomic-based methods (utilising DNA information). Pedigree-derived heritability estimates are based on the resemblance of relatives (Visscher *et al.*, 2008) and are more widely available for use due to lower costs since genotyping is not required.

Canine behavioural genetics is a growing field of interest within the scientific community (Rigterink and Houpt, 2014). As behavioural variation is impacted by genetics, environment and experience (Mackenzie *et al.*, 1986; Åkerberg *et al.*, 2012; Friedrich *et al.*, 2019), it can be challenging to describe the true phenotypic behaviour of animals, including dogs (Overall *et al.*, 2014). It is essential that behaviour is described objectively, as a common issue amongst behavioural studies is the use of varying, ill-defined and contrasting terminology. To build on the work of others, and to expand knowledge on specific behaviours and traits, consistency in recording and descriptions of what behaviour(s) have been measured is required (Bellamy *et al.*, 2018).

For behaviour, it is common to report genetic variance and heritability estimates as broader traits rather than specific individual behaviours (Ilska *et al.*, 2017) due to the complexity and overlapping of behaviours displayed by animals. Behavioural heritabilities in dogs range from 0.0 (Wiener *et al.*, 2015) to 0.8 (Liinamo *et al.*, 2007), however the majority of estimates are reported to be under 0.4 (Ilska *et al.*, 2017). It is suggested

this may potentially be due to the large effects of environment, learning and training on behavioural displays (Mackenzie *et al.*, 1986). Hradecká *et al.* (2015) undertook a meta-analysis on the heritability of behavioural traits reported from 48 studies and extracted five higher categories (Environment, Herding, Hunting, Play and Physical characteristics) which were reviewed for further analysis. In general, low heritabilities (h^2 average ranged from 0.09 to 0.15) were reported for the traits with differences between breed groups reported.

The heritability of behavioural traits in dogs measured using owner-reported questionnaires have been studied (Friedrich *et al.*, 2019; MacLean *et al.*, 2019). Heritabilities ranging from 0.27 to 0.73 have been reported for C-BARQ data from more than 14000 dogs from over 100 breeds (MacLean *et al.*, 2019), with variation attributable to genetic differences. Friedrich *et al.* (2019) also reviewed data for around 1000 German Shepherd Dogs (GSD) using C-BARQ questionnaire responses and reported heritability estimates to range from 0.00 to 0.23. Schiefelbein (2012) found mostly low heritabilities for C-BARQ behavioural traits in two North American populations of guide dogs; results ranged from 0.01 to 0.46 for approximately 3000 Labrador Retrievers, Golden Retrievers and GSD at six and 12 months of age. Heritabilities are calculated for individual populations, but differences found for similar populations highlights the importance of not grouping populations that vary in training practices, qualification criteria and breeding animals for heritability studies.

Whilst behaviour questionnaires have limitations (such as owner bias, recall bias and assessor variance, see Chapter 1), collecting data in this way enables data for large numbers of individuals and groups to be reviewed which is a requirement for heritability studies (Liinamo *et al.*, 2007). Questionnaires also enable fast and efficient data collection when compared to practical assessments (Hall and Wynne, 2012). Questionnaires have been shown to be reliable for collecting data on behavioural traits in dogs, which can provide the information required for

genetic studies where hundreds or thousands of data points are needed (Wiener and Haskell, 2016). As heritability estimates are population-specific, and can vary across time, environments and between populations (Schiefelbein, 2012; MacLean *et al.*, 2019), it is important to continually review and monitor data for different populations.

There is an increase in the demand and prevalence of working and assistance dogs worldwide (Chen *et al.*, 2021a). Therefore, understanding the heritability of behaviours can help to improve breeding programmes and production of working and assistance dogs. The heritability of behaviour in guide dogs was first reported by Goddard and Beilharz (1983) for 249 Labrador Retrievers in Australia, with heritabilities ranging from 0 (nose distraction and hearing sensitivity) to 0.58 (nervousness). Since this study, researchers have reported behaviour, traits and qualities that are linked to success in working roles such as guide dogs (Serpell and Hsu, 2001; Arata *et al.*, 2010; Kobayashi *et al.*, 2013; Harvey *et al.*, 2017; Hunt *et al.*, 2020) and more recently the heritability of the behaviours (Evans *et al.*, 2015; Chen *et al.*, 2021a).

Breeding programmes using phenotype selection alone will find it problematic to maintain and improve performance over many generations (Chen *et al.*, 2021a). Once heritability estimates have been calculated, estimated breeding values (EBVs) can be determined for individuals, which provide a more powerful and robust method for selection compared to phenotype selection alone. Historically GDUK have explored calculating behavioural EBVs from Canine Assessments Summaries (CAS) completed by puppy raising and training staff. However the subjective terminology used for these measures influenced the ability to reliably calculate EBVs, and few CAS scores were shown to be heritable (Evans, 2014). GDUK would benefit from the use of EBVs for behaviour should it be possible to calculate them from data collected through the r-PWQ2.

3.2 Aims and Objectives

3.2.1 Research question

Are data captured through behavioural questionnaires heritable in a population of guide dog puppies?

3.2.2 Aim

To review the heritability of behavioural data collected using the r-PWQ2 at 5M, 8M and 12M.

3.2.3 Objectives

- To undertake heritability analysis for r-PWQ2 behavioural component scores for puppies collected at five, eight and 12 months of age
- To determine whether estimated breeding values (EBVs) can be calculated and implemented within GDUK' breeding programme using the r-PWQ2.

3.3 Materials and methods

3.3.1 Study design and setting

This study followed the same study design and methods as described in Chapter 2 (see 2.3.1 and 2.3.2). A mixture of both retrospective and prospective methods were used to collect behavioural questionnaires for GDUK puppies born over four years (2016-2019) that were subsequently placed into puppy raising. Data were gathered for puppies at five (5M), eight (8M) and 12 months (12M) using a refined version of the refined r-PWQ (Hunt *et al.*, 2020), the r-PWQ2 (Chapter 2).

3.3.2 Study animals

Subjects were GDUK puppies born between 07/08/2016 and 18/07/2019 that were placed into puppy raising. This study was undertaken using questionnaire data for Labradors, Golden Retrievers and their crosses (n=2757; 1311M/1264F). Data were available for 492 dogs at 5M, 2063 at 8M and 1374 at 12M. Dogs were born between 2016 and 2019 (2016 n=284, 2017 n=894, 2018 n=720, 2019 n=677) and were from 470 litters produced by 97 sires and 318 dams.

German Shepherd Dogs (n=146), Golden Retriever x German Shepherd Dogs (n=107) Curly Coated Retrievers (n=2), Curly Coated Retriever crosses (n=5), and Standard Poodle crosses (n=36) were removed from the data set. Three Golden Retriever x Flat Coated Retriever dams were also removed, accounting for 21 puppies.

3.3.3 Variables and data sources

3.3.3.1 *Phenotype data*

Phenotype data were collected using the r-PWQ2 behavioural questionnaire as previously described in Chapter 2. Data were collected from 39 questions which produced 10 behavioural component scores, in addition to information about the puppy raisers home environment.

Information was also collected from GDUK's electronic database (GDIR) including date of birth, sex, breed, dam ID, dam name, sire ID, sire name and litter ID. Puppy raiser ID, PDA, puppy raising region and team were also obtained from GDIR. Data collected at five, eight and 12 months of age were used for analyses.

The proportion of Labrador for each dog was calculated based on three previous generations to include parents, grandparents and great grandparents (Figure 3.1 and 3.2). Labradors were given a score of 1 and

Golden Retrievers a score of 0, with proportions varying from 0.06 to 0.75.
For example:

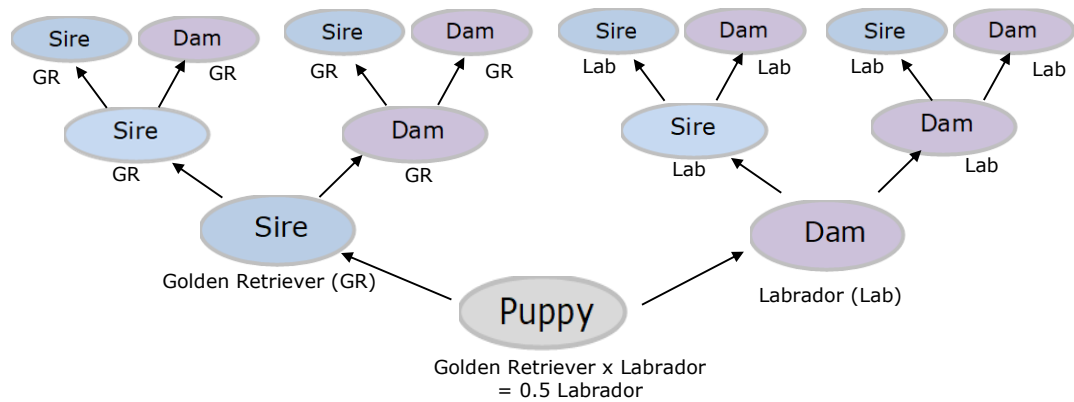


Figure 3.1 Breed calculation for 0.5 Labrador.

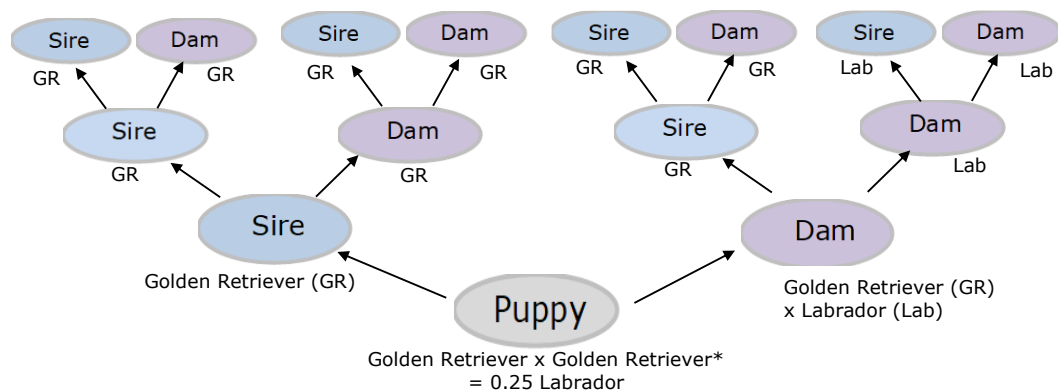


Figure 3.2 Breed calculation for 0.25 Labrador. Breeds are reported sire first with an x indicating a cross breed and * indicating the dam was also a crossbreed.

3.3.3.2 Pedigree data

An existing pedigree file that contained all dogs represented in the phenotype was used and recoded for analysis (pedigree recoded by Dr Tom Lewis). Within the pedigree file, dogs were assigned a unique identifier (*pid*). The *pid* for each individual dog, their sire and their dam were collated in one data file for analyses. A phenotype file was produced containing all dog information (*pid*, dog name, date of birth, year of birth, breed, sex, sire ID, dam ID and litter ID), component data and other confounding variable data collected in the additional information (presence

of children, dogs and animals, number of adults, attendance at puppy classes or the workplace) and data held within GDIR (puppy raiser ID, PDA, puppy raising region, number of puppies previously raised by the puppy raiser and neutering date). Puppy raising region was not included in analyses as the factor was nested in both puppy raiser ID and PDA.

3.3.4 Bias and study size

Biases for the heritability analyses were the same as the those in data collected in Chapter 2 (see 2.3.5). Sample size was smaller for the present analysis due to the removal of some breeds (see 3.3.2). GDUK have smaller numbers of Standard Poodle and German Shepherd Dog breeding stock, these breeds were removed from heritability analyses.

3.3.5 Quantitative variables and statistical methods

3.3.5.1 *Variables*

Following the factor analyses undertaken in Chapter 2, data were collated for this cohort of dog. Data were grouped in the same way as described in 2.3.6.3. All missing responses were coded as NA.

3.3.5.2 *Heritability*

Heritability analyses were undertaken on component data for Labradors, Golden Retrievers and their crosses (n=2575). Residual maximum likelihood (REML) is the standard method for estimating genetic variance components in complex pedigrees (Visscher et al., 2008). A univariate linear mixed model using ASReml (Windows version 4, VSN, International Ltd) was run for each behaviour component at each time point to estimate the heritability. Separate models were run for 5M, 8M and 12M data sets. To determine the most appropriate model to run at each time point, the log likelihood was compared for different models when the different random effects were included or excluded. Every model included pid as a random

effect, with all possible combinations of the three other factors (litter ID, PDA and puppy raiser ID) compared. One model at each time point was compared using likelihood ratios when including or excluding combinations of each of the random effects (litter, Puppy Development Advisor and Puppy Raiser ID). The animal effect was included in each model (see Appendix 8.6 for log likelihood comparisons between models). Puppy raiser ID was not included as a random effect due to the high impact on the residual variance in eight- and 12-month data sets. Kinship analyses (data not reported) revealed that when puppy raisers raised more than one puppy in the data set, there were relationships apparent between puppies, so puppy raiser ID was not sufficiently random. Kinship analyses were calculated from the pedigree using widely available genetic algorithms e.g. Meuwissen and Luo (1992) and Kirkpatrick *et al.* (2018). Puppy raiser ID was also nested within puppy raiser experience which was included as a fixed effect in the model.

Heritability models were run in ASReml (Windows version 4, VSN, International Ltd) for pedigree-based estimates by fitting the following model:

$$y = 1\mu + X\beta + Z\alpha + \varepsilon$$

where y is a vector of behaviour traits, μ is the overall mean, β is a vector of fixed effects with X as the corresponding incidence matrix, Z is the incidence matrix for the random additive polygenic effect, α is a vector of random additive polygenic effects distributed as for the pedigree-based estimates and ε is a vector of residual errors.

Fixed effects included in all models were the sex and breed (proportion of Labrador) of the individual as fundamental biological factors of importance. At 5M, additional fixed factors included in the model were puppy raiser experience and the presence of children. At 8M, additional fixed factors included in the model were, puppy raiser experience, presence of children

and location of the puppy raisers home. At 12M, additional fixed factors included in the model were puppy raiser experience, presence of children, number of adults in the home and the age at neutering in days. Fixed effect results were reported for the 2757 Labradors, Golden Retrievers and their crosses. The significance of fixed effects were determined by dividing the Z-ratio variance by the standard error. Values greater than 1.96 were considered significant. The effect size $\pm$ standard errors are reported for significant factors, where the effect size is an increase or decrease in scale units (using the 0-100 scale for r-PWQ2 scores).

3.4 Results

The final pedigree file contained 59,097 individuals, of which 30,766 were female and 28,331 were male. The most common breeds represented in the pedigree were Labrador Retrievers (n=22,055), Golden Retriever x Labrador or Labrador x Golden Retriever (n=15,056) Golden Retrievers (n=8261) and German Shepherd Dogs (4786). There were 4557 unique sires and 8068 unique dams. The mean number of offspring per sire was 11.8, with a minimum of 1 and a maximum of 621. The mean number of offspring per dam was 6.6, with a minimum of 1 and a maximum of 61.

The data set for Labrador, Golden Retrievers and crosses of these two breeds with 5M, 8M and 12M data contained a total of 2575 individual. The mean number of offspring per sire was 26.5 and per dam 8.1. The average number of dogs per litter were 5.5. Dogs were from 11 breeds (Table 3.1).

Table 3.1 Puppy breeds represented in r-PWQ2 heritability analyses, where breed is shown as a proportion of Labrador for 5M, 8M and 12M data sets.

Proportion of Labrador	Puppy breed(s)	Total sample (n)	5M (n)	8M (n)	12M (n)
0.0000	Golden Retriever	316	100	248	125
0.0625	Golden Retriever x Golden Retriever*	5	5	3	0

0.1250	Golden Retriever x Golden Retriever*	118	38	84	53
0.2500	Golden Retriever x Golden Retriever* and Golden Retriever x Labrador*	128	17	113	61
0.5000	Golden Retriever x Labrador and Labrador x Golden Retriever	1036	155	793	625
0.6250	Labrador x Golden Retriever*	16	0	16	12
0.7500	Labrador x Golden Retriever* and Labrador x Labrador*	261	21	221	169
1.0000	Labrador	695	156	584	301
TOTAL		2575	492	2063	1374

An * indicates the dam was a crossbreed.

3.4.1 5M heritability

5M heritability models were run for each r-PWQ2 component with the fixed effects of sex, proportion Labrador, puppy raiser experience and the presence of children. Heritability at 5M ranged from 0.00 (C6 'Impulsivity') to 0.29 (C4 'Anxiety'; Table 3.2). Four components had a heritability estimate above 0.1, suggesting that over 10% of the component variation can be accounted for by additive genetic variation. C4 demonstrated significant heritability, with all other components not determinably larger than zero.

Table 3.2 5M heritability estimates for r-PWQ2 components. Heritability estimates (h^2) and their standard errors (S.E.) are shown to two decimal places.

Component	5M r-PWQ2 component	h^2	$\pm$ S.E.
C1	Distraction and Human Focus	0.04	0.07
C2	Separation	0.04	0.08
C3	Anxiety	0.29*	0.13
C4	General Excitement	0.13	0.09
C5	Attachment and Attention Seeking	0.00	0.09
C6	Impulsivity	0.00	0.06

C7	Energy	0.13	0.09
C8	Anticipatory Excitement	0.00	0.07
C9	Body Sensitivity	0.13	0.10
C10	Focus	0.04	0.09

* indicates a significant heritability.

Fixed effect results following heritability analyses were reported for Labradors, Golden Retrievers and their crosses. Factors that impacted component scores at 5M included children and puppy raiser experience. (see Table 8.21 in Appendix 8.7 for breakdown). C1 (Distraction and Human Focus) was significantly impacted by the presence of children (more desirable scores for puppies that did not live with children; effect 4.99 ± 2.11) and experience of puppy raisers (more desirable scores for puppy raisers that had raised 5-9 puppies (effect -7.49 ± 2.33) and >20 puppies; effect -11.19 ± 4.87) when compared to other puppy raiser experience groups. C6 (Impulsivity; effect 7.05 ± 2.73) and C10 (Focus; effect 5.15 ± 1.66) were impacted by the presence of children, with more desirable scores present for puppies that did not live with children. C6 was also impacted by the experience of puppy raisers that had raised between five and nine puppies (effect -7.69 ± 3.00), with more desirable scores seen by this group and those raising >20 puppies. C5 (Attachment and Attention Seeking) was impacted by puppy raiser experience (more desirable scores for puppy raisers that had raised >20 puppies; effect -12.45 ± 4.62) when compared to the other levels of experience.

3.4.2 8M heritability

The 8M heritability models were run for each r-PWQ2 component with the fixed factors of sex, proportion Labrador, puppy raiser experience, presence of children and location of the puppy raisers home. Heritability at 8M ranged from 0.00 (C10; Attention seeking) to 0.09 (C5; Anticipatory Excitement; Table 3.3). No components had a heritability estimate above 0.1, suggesting that less than 10% of all component variation was accounted for by genetics at 8M of age. C1 and C5 demonstrated

significant heritabilities, with all other components not determinably larger than zero.

Table 3.3 8M heritability estimates for r-PWQ2 components. Heritability estimates (h^2) and their standard errors (S.E.) are shown to two decimal places.

Component	8M r-PWQ2 component	h^2	$\pm$ S.E.
C1	Impulsivity	0.08*	0.04
C2	Anxiety	0.02	0.03
C3	Separation	0.05	0.04
C4	Distraction	0.02	0.03
C5	Anticipatory Excitement	0.09*	0.04
C6	Attachment	0.05	0.04
C7	Energy	0.04	0.03
C8	Focus	0.03	0.03
C9	Body Sensitivity	0.03	0.03
C10	Attention Seeking	0.00	0.03

* indicates a significant heritability.

Fixed effect results following heritability analyses were reported for Labradors, Golden Retrievers and their crosses. Factors that impacted component scores at 8M included presence of children, puppy raiser experience, location of the home, proportion of Labrador and sex (see Table 8.22 in Appendix 8.7 for breakdown). C1 (Impulsivity) was significantly impacted by more inexperienced puppy raisers (raising less than five puppies; effect 6.68 ± 1.59) and highly experienced puppy raisers (raising more than 20 puppies; effect -7.91 ± 3.55). Puppies raised by highly experienced raisers had more desirable scores. For C1, puppies raised in an urban location had less desirable scores when compared to the other locations (effect 4.97 ± 1.81). C2 (Anxiety) was impacted by the presence of children (more desirable scores when not living with a child; effect 1.78 ± 0.89) and the proportion of Labrador (less desirable scores seen for pure Labradors and 12.5% Labradors; effect -6.35 ± 1.68). C3 (Separation) was impacted by the puppy raiser home being in a suburban (effect 1.57 ± 0.79) or urban location (effect 2.92 ± 1.09) when compared to rural locations, with more desirable scores seen for puppies raised in rural locations. C6 (Attachment) was impacted by the presence of children (more desirable scores when no children were in the home; effect 3.06

± 1.47), the sex of the puppy (more desirable scores for females; effect 2.83 ± 1.32), the experience of puppy raiser (more desirable scores when raising >20 puppies; effect -8.60 ± 4.02) and the presence of the home in an urban location (more desirable scores for puppies raised in urban locations; effect 6.82 ± 2.05). C7 was impacted by the proportion of Labrador (most desirable scores for 6.25 and 25% Labradors; effect 5.06 ± 2.52), presence of children (more desirable scores when children present; effect -5.42 ± 1.10) and sex (more desirable scores for males; effect -3.32 ± 0.99). C8 (Focus; effect 3.26 ± 1.40), C9 (Body Sensitivity; effect 3.15 ± 1.59) and C10 (Attention Seeking; effect 3.44 ± 1.53) were influenced by home location, with less desirable scores seen when raised in urban environments. C10 was also impacted by the presence of children (more desirable scores seen when no children present; effect 3.52 ± 3.44).

3.4.3 12M heritability

12M heritability models were run for each r-PWQ2 component with the fixed factors of sex, proportion Labrador, puppy raiser experience, presence of children, number of adults in the home and the age at neutering in days. Heritability at 12M ranged from 0.00 (C4 'Impulsivity' and C5 'Focus') to 0.08 (C2 'Anxiety', C6 'General Excitement' and C9 'Body Sensitivity'; Table 3.4). No components had a heritability estimate above 0.1, suggesting that less than 10% of all variation in components can be accounted for by genetics at 12M of age. No component had a significant heritability determinably larger than zero

Table 3.4 12M heritability estimates for r-PWQ2 components. Heritability estimates (h^2) and their standard errors (S.E.) are shown to two decimal places.

Component	12M r-PWQ2 component	h^2	$\pm$ S.E.
C1	Separation	0.06	0.05
C2	Anxiety	0.08	0.05
C3	Impulsivity	0.00	0.00
C4	Focus	0.00	0.00

C5	Attachment and Attention Seeking	0.03	0.04
C6	General Excitement	0.08	0.06
C7	Anticipatory Excitement	0.07	0.05
C8	Energy	0.07	0.05
C9	Body Sensitivity	0.08	0.05
C10	Human Focus	0.04	0.04

Fixed effect results following heritability analyses were reported for Labradors, Golden Retrievers and their crosses. Factors that impacted component scores at 12M included presence of children, puppy raiser experience, proportion of Labrador, sex, number of adults in the home and date at neutering (see Table 8.23 in Appendix 8.7 for breakdown). The presence of children in the home impacted C1 (Separation; effect 2.10 ± 0.78), C2 (Anxiety; effect 3.17 ± 1.10), C5 (Attachment and Attention Seeking; effect 3.16 ± 1.41), C7 (Anticipatory Excitement; effect 2.75 ± 1.30) and C8 (Energy; effect -3.47 ± 1.40). Four components showed more desirable scores when no children lived in the home, with C8 showing more desirable scores in the presence of children. Puppy raisers with less experience (<5 puppies) were significantly different to the other puppy raiser experience groups for C3 (Impulsivity; effect 4.04 ± 1.79) and C4 (Focus, effect 3.60 ± 1.47), with puppies showing less desirable scores. The proportion of Labrador influenced C2 (Anxiety; effect -6.46 ± 2.85), with less desirable scores seen for pure Labradors, and sex impacted C4 (Focus; effect 4.68 ± 1.87), with more desirable scores for females. The number of adults in the home impacted C7 (Anticipatory Excitement; 1 adult effect -12.18 ± 5.41 ; two adults effect -12.84 ± 5.30 and four or more adults effect -11.86 ± 5.89) and two adults in the home impacted C10 (Human Focus; effect -9.72 ± 4.10), with least desirable scores seen when five or more adults were in the home, with the most desirable scores found when puppies lived with two adults. Neutering status impacted C2 (Anxiety; effect 0.02 ± 0.01), C3 (Impulsivity; effect 0.03 ± 0.01), C4 (Focus; effect 0.03 ± 0.01), C5 (Attachment and Attention Seeking; effect 0.04 ± 0.01) and C7 (Anticipatory Excitement; effect 0.02 ± 0.01), with more

desirable scores seen for puppies that were spayed after the questionnaire window (i.e. after 12.5months).

3.4.4 Summary of heritability results

The questions included within each component differed slightly between the three time points based on results from the PCA analysis (see Chapter 2), therefore not all components are directly comparable. However, what is apparent from the data is that heritabilities at 5M were generally the highest for the population reviewed. The component with the highest heritability at 5M was C3 'Anxiety' at 0.29, at 8M was C5 'Anticipatory Excitement' at 0.09 and at 12M was C3 ('Anxiety'), C6 ('General Excitement') and C9 ('Body Sensitivity'), all at 0.08. The highest heritabilities in general appear to be for the 'Anxiety' and 'Excitement' related traits. There were also fewer fixed effects that impacted results at 5M than compared to 8M and 12M.

3.5 Discussion

The present chapter met its aims by calculating the heritability for r-PWQ2 behavioural component scores for puppies at 5M, 8M and 12M.

Heritabilities for questionnaire components for puppies at 5M were higher than those at 8M and 12M. The highest heritability at 5M was 0.29 for Anxiety, whereas heritability estimates at 8M and 12M were all below 0.1 indicating that less than ten percent of the behavioural variation present in the phenotype measured using the r-PWQ2 could be accounted for by genetics. One component at 5M (Anxiety) and two components at 8M (Impulsivity and Anticipatory Excitement) demonstrated significant heritabilities. No components demonstrated significant heritabilities at 12M.

The success of any breeding program incorporating behavioural measures depends on the understanding of the genetics behind the behaviours. Within the field of canine behaviour genetics, heritabilities are commonly reported to be below 0.4 due to the large impact of environment on behaviour (Ilska *et al.*, 2017). A review by Hradecká *et al.* (2015) reported that of the 48 studies examined, heritabilities for behavioural traits were low (h^2 average ranged from 0.09 to 0.15). The present study found questionnaires collected at 5M to show generally higher heritabilities than questionnaires collected at 8M and 12M. Friedrich *et al.* (2019) reported heritabilities for components based on C-BARQ questions in a population of 1041 German Shepherd Dogs. Excitability heritability was 0.05, Separation was 0.00, Attachment/Attention-seeking was 0.00, Non-social fear was 0.12 and Touch sensitivity 0.02. The low heritabilities reported in the present study for most components, except for Anxiety at 5M, are therefore consistent with behaviour heritabilities reported in dogs by others.

The dogs included at the three different time points were not directly comparable in the present study. The 5M data set was smaller than 8M and 12M due to GDUK collecting the r-PWQ as standard at 8M and 12M. Nearly half of the dogs in the present 8M and 12M sample ($n=1266$, 48.9%) had questionnaires in both the 8M and 12M data sets. Similarly, nearly half ($n=247$, 44.8%) of dogs from the 5M sample had 8M data, but only one dog with 5M data also had data at 12M. As discussed in Chapter 2, data collection was impacted by the Covid-19 pandemic. However, for this population, many of the dog's common ancestry and were linked by the pedigree and the dogs were born over the span of four years. Therefore, the difference in the groups of dogs should not have impacted heritability analyses. A future study to review the heritability behavioural questionnaires collected for the same dogs at the three time points would be required to confirm this suggestion.

MacLean *et al.* (2019) reported that measures between breeds yield higher heritability estimates, averaging five times higher than within breed estimates, due to wider genetic variation seen between breeds. It is therefore common for studies to be breed specific. Schiefelbein (2012) reported considerably lower heritabilities at 12 months in German Shepherds ($h^2=0.04$) than compared to Golden Retrievers ($h^2=0.27$) for C-BARQ 'Non-Social Fear', suggesting there could be a phenotypic breed difference or that genetics accounts for more variation in one breed than another. Within this study it was not possible to include German Shepherd Dogs or Standard Poodle crosses due to low numbers. This is a limitation of the study in that not all breeds within the GDUK breeding population have been reviewed. It will be many years before enough data are available to review heritability for the less common breeds within the breeding programme due to the number of animals needed for accurate analyses. This is potentially a risk to the breeding programme, as behaviour cannot be accurately accounted for within these breeds when selecting for behavioural traits.

This study found the impact of puppy raiser to be high for all questionnaire time points. This was to be expected due to similar findings within the larger population reported in Chapter 2, to the type of data being assessed and the variation seen between individuals when reporting and recording dog behaviour. Harvey (2014) previously showed the inter and intra-rater reliability of the Puppy Walker Questionnaire (PWQ) to be acceptable. During heritability analyses when adding puppy raiser ID as a random effect, the heritability values increased. This result was opposite to the expected outcome when adding another random effect to the model. When investigating the potential reasons for this, the kinship of dogs and puppy raiser ID was reviewed (results not reported). It became apparent that puppy raisers commonly raised puppies from certain lineages or sires/dams. Therefore, when reviewing heritability, puppy raiser ID was not a completely random effect. GDUK should consider whether requests for puppy raisers to raise certain lines or certain puppies should be agreed

due to the later impacts on heritability analyses. However, it may be that puppies are more successful in the guide dog training programme when raised by a similar puppy raiser.

The heritability of guide dog behaviour has been explored by several researchers. The highest behavioural heritabilities reported for the GDUK population to date have been when using practical assessments in puppies ($h^2=0.02-0.25$; Evans *et al.*, 2015). Evans (2014) reviewed the heritability of subjective behaviour assessments (CAS), as part of a large genetic evaluation of GDUK. Heritabilities were reported to range from 0.02 (obedience at eight months of age) to 0.16 (body sensitivity at the first CAS assessment in advanced training) for Labrador Retrievers. Interestingly Evans reported no heritability estimates to be detectably larger than zero for CAS elements at the first assessment in puppy raising. Also, more heritability estimates were detectably larger than zero for CAS elements at eight months of age (9/43) than at five months (7/43) or 12 months (4/43) for Labrador Retrievers. The present study reported somewhat different findings as more components at five months showed higher heritabilities when compared to eight and 12 months. Evans (2014) reported differences in heritability estimates for Labradors when compared to Golden Retrievers and crosses of the two breeds. Heritability estimates for Golden Retrievers ranged from 0.07 (obedience - first assessment in early training) to 0.25 (eagerness - first assessment in advanced training). For Golden Retriever and Labrador crosses, heritability estimates that were detectably larger than zero, ranged from 0.01 (behaviour on transport at five months, body sensitivity at five, eight and 12 months, calmness at the first assessment in puppy walking, eagerness at eight months, obedience at eight months and stress resilience at eight months) to 0.15 (distraction - first assessment in advanced training). The present study ran one model accounting for the proportion of Labrador of the individual. This was due to the number of individuals in the data set, and there not being enough data to run separate purebred and cross breed models. The results from the present study would therefore be more

comparable to the crossbreed model by Evans (2104) due to the mix in breeds in the data set (Golden Retrievers n=316, Labradors n=695, crossbreeds n=1564). It would be interesting for future studies using the r-PWQ2 to review differences between pure and crossbreeds used by GDUK once data from a larger number of dogs are available for analysis.

There are differences and similarities in the reported behavioural heritability between the r-PWQ2 and CAS for guide dog behaviour. It is difficult to directly compare the findings from Evans (2014) to the components in the present study due to differences in the terminology and data collection methods used, and the time at data collection. CAS reports were not collected at standardised time points, many were dependent on the dogs' speed of progression and staff member preferences for completing reports. During the Epidemiology of Guide Dog study (Asher *et al.*, 2016) considerable variation was found for the number and timing of CAS reports for each dog. Therefore, comparing two dog scores is not necessarily reflective of the same dog stage and age. The higher heritabilities reported by Evans (2014) were for CAS measures in advanced training, when the dogs would typically be between 18 and 24 months of age. CAS reports have previously been reported to be heavily influenced by assessor and the measures use subjective language (Asher *et al.*, 2016). Due to the subjective measures and the way the data were collected, the data were not robust or reliable enough to use for creating behaviour estimated breeding values (EBVs). GDUK has therefore never used behavioural EBVs within the breeding programme (M. Bottomley 2020, *personal communication*, 20 July 2020).

Over the last 10 years, GDUK has invested heavily in updating the way behaviour is measured across the organisation, including the implementation of the r-PWQ (Hunt *et al.*, 2020) and removal of subjective behaviour descriptors from CAS reports such as "confident" and "willing". Using a more reliable and valid measure of behaviour (whilst acknowledging potential data biases from using a questionnaire) have

enabled the present study to review the heritability of behaviour within guide dog puppies between five and 12 months of age. The questionnaire was designed to capture behaviours related to the likelihood of succeeding in the guiding role, including both working and non-working related behaviours. Non-working behaviours, also referred to as social behaviours, have been reported by guide dog owners as equally as important to working behaviours (York and Whiteside, 2018) so should be considered when looking at the behaviour of potential guide dogs.

Anxieties are reported as one of the greatest behaviour problems within dogs (Salonen *et al.*, 2020). Anxiety-related behaviours are common reasons for guide dogs being withdrawn from the programme (Goddard and Beilharz, 1982; Serpell and Hsu, 2001; Moxon, 2018) and also for the early retirement of guide dogs (Caron-Lormier *et al.*, 2016). The present study found that the 'Anxiety' component at 5M had the highest heritability (0.29) amongst all components measured. The heritability for the comparable component was found to be lower at the 8M ($h^2=0.02$) and 12M ($h^2=0.08$). Schiefelbein (2012) found the heritability of a similar trait, C-BARQ Non-social fear, to be 0.27 for Golden Retrievers at 12 months and Goddard and Beilharz (1983) found 'nervousness' to have a heritability of 0.58. Harvey (2014) reported 5.3% of the variation in General Anxiety within the PWQ to be due to sire and litter effects. GDUK has monitored Anxiety through behaviour withdrawal incidences due to reasons such as 'high suspicion' and 'low stress resilience'. When reviewing the breeding alerts used by GDUK in 2020, the withdrawal average for 'Suspicion' for all Labradors and Labrador cross Golden Retrievers was 9% and slightly lower at 7% for Golden Retrievers (unpublished GDUK data, 2020). When reviewing 'stress resilience' as a CAS behaviour measure, Evans (2014) reported the highest heritability to be 0.04, which was for the first assessment on entry to training. Similarly, 'suspicion' heritability results were 0.04 at five months of age, and 0.06 during the first advanced training assessment. It could be argued that to date GDUK have not been accounting for the genetic influence of the

behaviours but have selected based on phenotype alone. The breeding programme could be improved if behaviours with a heritable basis are used for selection to breed towards/away from certain traits, and that additional efficient could come from not selecting on traits with no detectable heritability. The significant heritability of 0.29 at 5M reported in the present study suggests that it may be possible to use the component to aid in breeding selection at GDUK.

In the present study, the Excitement component showed a heritability of 0.13 at 5M, compared with 0.09 at 8M and 0.08 at 12M. Excitability has been described as a reason for rejection from the guide dog programme in the USA (Serpell and Hsu, 2001). The questions asked in the component include the degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at the home, the ability to control the dog, presence of lunging, trouble settling down and the ability to remain calm. Saetre *et al.* (2006) estimated the heritability of greeting behaviour in a practical assessment (the Swedish Dog Mentality Assessment) to be 0.05 in 5964 German Shepherds and 0.09 in 4589 Rottweilers. During development of the PWQ, Excitement was shown to be impacted by sire (11.2% of the variation) and litter (10.1% of the variation) (Harvey, 2014). GDUK have historically monitored extremes of excitability by alerting for withdrawal reasons related to 'hyperactivity' or a dog being 'boisterous', accounting for 4% of Labradors, Golden Retrievers and their crosses (unpublished GDUK data). GDUK may have been selecting for these behaviours and due to the heritability at 8M showing significance, it is suggested GDUK may be able to select furthermore in the future.

The Energy component also had a 5M heritability of 0.13 in the present study, with heritabilities lower at 8M (0.04) and 12M (0.07). The energy component asks questions related to playfulness and activity levels. Hradecká *et al.* (2015) found the heritability of play to be on average 0.09 during a meta-analysis, similar to the results in the present study. High energy levels in working guide dogs would be an undesirable behaviour

and could cause a dog to be removed from the programme. The GDUK breeding programme have previously monitored 'Hyperactive' or 'Boisterous' behaviour collectively within a higher category of 'Antisocial behaviour', accounting for 4% of the removals within breeds. Harvey (2014) reported the Energy trait to have the highest sire impact when compared to all other traits, accounting for 12.0% of the variation, and Evans (2014) reported CAS heritability for 'calmness' (the opposite of an energetic dog) to range from 0.08 at five months to 0.11 for the first assessment in early training. Results from the present study suggest it may be possible to influence future behaviours through breeding, but the heritability estimates were not significant, and therefore the influence of environmental factors should also be explored.

Another component that showed low heritability, but above 0.1 was Body Sensitivity at 5M, with a heritability of 0.13. The same component was less heritable at 8M (0.03) and 12M (0.08). If a dog were to display body sensitivity behaviours, the behaviour would impact the success of a guide dog due to the necessity for a harness to be worn when guiding the client. This behaviour is therefore more important for some populations than others, such as pet dogs, where displaying the behaviours may not be so problematic. The questions asked within this component specifically relate to being handled and being groomed. The highest heritabilities within the PPA assessment for over 2100 Golden Retriever and Labrador pure breeds and crosses were related to human contact and interaction with a human (following and response to stroking $h^2=0.23$, response to the assessor $h^2=0.19$) (Evans *et al.*, 2015). Sire variation was noted to account for 5.8% of the variation in 'Body Sensitivity' within the PWQ (Harvey, 2014) and Evans (2014) reported Body Sensitivity heritability to be 0.10 for the first assessment in early training and 0.16 for the first assessment in advanced training. Other guide dog schools have reported heritabilities for harness sensitivity (defined as dog dropping rear quarters when harness handle lays on back) to be 0.38 (Riser *et al.*, 2021), 0.47 during the 'training test' (Chen *et al.*, 2021a) using the Behaviour Checklist

(IWDBA, 2025) and 0.33 for body sensitivity (defined as strong response to touch and leash corrections; Goddard and Beilharz (1983). In contrast, others have described low touch and body sensitivity related heritabilities such as Friedrich *et al.* (2019) and Schiefelbein (2012). It is suggested that it may be possible to select for Body Sensitivity behaviours in this guide dog population and other guide dog populations. However, the heritability estimates were not significant and therefore the influence of environmental factors should also be explored. The r-PWQ2 component contains only two questions, so the addition of more questions related to body sensitivity could also be considered to ensure a wider range of behaviours related to this component are being captured.

Some components showed low heritabilities at all three time points, suggesting that these behaviours generally have a greater variation due to environmental influences. This highlights the importance of adequate training and socialisation for puppies. GDUK invests heavily in early puppy socialisation whilst puppies are in the nest, due the impact that reduced socialisation has on the outcome and behaviour of future guide dogs (Vaterlaws-Whiteside and Hartmann, 2017). More recently, GDUK has rolled out PREP (Preparing Puppies for Excellent Partnerships) to standardise and support puppies and volunteers during puppy raising. It is unknown how the impact of the new raising practices will affect the behaviours seen within puppies in the organisation. Future heritability studies will be able to compare data from this population to future populations.

The use of EBVs can reduce the impact of missing data in an individual as the data for related individuals is taken into account (O'Neill and Packer, 2016). The completion rates for the r-PWQ2 range from between 70 and 80% (unpublished data) for the population at the three different time points. Therefore, if GDUK were to rely on phenotype alone, up to one fifth of the population would have data missing. Another benefit of an EBV, is the ability to monitor and manage different traits simultaneously, which

reduces the risk of inadvertently selecting other traits within the population (O'Neill and Packer, 2016). GDUK currently does not utilise behaviour EBVs. Findings from this study suggest that the r-PWQ2 can be used to develop EBVs for selected components (Anxiety, Excitement, Body Sensitivity and Energy) at five months of age, all of which have been behaviours linked to the success of guide dogs. Additionally, the results indicate that earlier measures could be more valuable as a selection criterion. Collecting data from early ages can enable scientists/researchers to understand the genetic impact of behaviour as there has been less opportunity for environmental variation and experiences between individuals.

The limitations of the present study have been discussed above and include different cohorts of dogs studied between the three time points, the removal of less popular breeds and the issue of validity using the r-PWQ2 (conversed in Chapter 2). Although the components suggestive of higher heritabilities at 5M were the components that demonstrated consistency between the three time points, and the groupings were not changed between the r-PWQ and r-PWQ2. It was not possible to control for breed within data collection, as this was based on the dogs bred by GDUK and the puppy raisers that completed questionnaires. A study with comparable numbers of pure and cross breeds would have enabled separate analysis to be undertaken to distinguish differences between the two groups. Future studies could ensure the same dogs are reviewed at all three time points, that data for the less popular breeds are analysed when enough data has been collected, and the analysis could be undertaken using the r-PWQ to compare heritabilities between the two questionnaire structures.

3.6 Conclusion

Determining the influence of genetics on behaviour is beneficial to breeding programmes as it may contribute to producing more animals behaviourally suited for the role. This study has demonstrated how a behavioural questionnaire completed by volunteer puppy raisers can provide heritability information to a guide dog breeding programme, with the research question answered.

Results suggest that behavioural components measured at 5M are more suitable for collecting heritable behaviour information, with the Anxiety component at 5M having the highest heritability ($h^2=0.29$). Comparisons of heritabilities from this study to previously reported heritabilities from subjective staff reports highlight an improvement in data collection within GDUK. Based on the results presented, GDUK could explore using EBVs for the four components with heritabilities above 0.1 at 5M. Results suggest that it would be possible for behavioural EBVs to be used by the breeding programme to influence behaviour in the population, and selection could be more accurate than using phenotype information. The organisation should also focus on ensuring r-PWQ2 completion rates are high, due to the importance of the data for the breeding programme. Should similar behaviours show heritabilities in other populations e.g. other assistance dogs/working dogs or pet dogs, it may be possible to influence behaviour by breeding. However, heritability analyses for different populations would need to be undertaken.

Chapter 4. Heritability of the behavioural responses of guide dog puppies to a series of problem-solving tasks

4.1 Introduction

Behavioural data are commonly measured by undertaking practical assessments or through behavioural questionnaires (Barnard *et al.*, 2016). Behavioural assessments based on direct observations enable researchers to collect in depth behavioural data for a series of tasks/tests, providing complementary information to data collected through questionnaires. Some studies report the heritability of behavioural assessments (Evans *et al.*, 2015; Bray *et al.*, 2021a; Chen *et al.*, 2021a), but due to the considerable resources required to carry out assessments for a large cohort, there appear to be less published studies when compared to questionnaire data. In previous chapters, this thesis has examined the use of behavioural questionnaires within the Guide Dogs UK (GDUK) training programme and shown the impact of environmental factors and assessor influences on the behaviours observed. Collecting behavioural observation data would enable another perspective to be considered for this population and would enable comparisons between heritability for observational data and questionnaire data.

Animal cognition is described as the mental capacities of nonhuman animals and the study of those capacities (Shettleworth, 2001). Assessments and concepts related to animal cognition are a growing area of focus, with dog cognition research increasing since 2000 (Arden *et al.*, 2016; Pinelli *et al.*, 2023). Published works report cognition in a range of species including mice (Hölter *et al.*, 2015), pigs (Held *et al.*, 2002), chimpanzees (Hopkins *et al.*, 2014) and fish (Salena *et al.*, 2021). The foundation for cognitive studies in animals began by exploring the ability for animals to use tools (Emery and Clayton, 2009). This field has since developed to include understanding how species can solve and interpret complex tasks.

Dogs are a useful model species to study cognition due to their sociability, trainability, intelligence, and ease around humans. Cognition in dogs was defined as “the study of the dog’s mind....including sensing, perceiving, learning, remembering, and reasoning” by Byosiére and Prasad-Shreckengast (2020). It could be suggested that cognition can be inferred from behaviour assessments where animals are required to process information/problem solve. Bräuer *et al.* (2020) criticised the term ‘animal cognition’ because the phrase is often applied with an anthropocentric viewpoint, with researchers focusing on cognitive skills that are human-like. There is also a difference in the motivation for problem solving between domesticated and wild animals (Thornton and Truskanov, 2022), so the term cognition could mean something different between species depending on the goals of the animal. Cauchoux *et al.* (2018) noted that behavioural and cognitive processes are important to facilitate how an individual interacts with the environment. This poses questions as to whether all behavioural assessments are therefore measuring cognition. Behavioural assessments are commonly conducted under strict standardised laboratory style conditions; thus, it can be questioned whether these are reflective of an animal’s typical environment. Another question therefore is, how do we differentiate between a behavioural assessment and a cognition assessment? Researchers make assumptions about the interpretation of the behavioural display and whether this may be an animal’s cognition, which is open to subjectiveness.

Cognitive traits have been reported to develop early in puppies and are important for success as working dogs (MacLean and Bray, 2019; Bray *et al.*, 2020; Foraita *et al.*, 2021). Undertaking behavioural assessments at a young age can reduce the impact of environmental effects on behavioural responses. For this reason, assessments have focussed on younger puppies, between six and 10 weeks of age, when experiences have primarily been in the nest. However, at this young age, puppies

experience rapid development starting from around two weeks of age to between 12-14 weeks of age (Freedman *et al.*, 1961). There can be a great difference between six and 10-week-old puppies, specifically in their capability to engage with a specific task. Alberghina *et al.* (2020) stated that seven weeks of age would be too early to reliably determine the temperament of a dog. Yet studies have successfully reported findings for puppies between six and seven weeks of age (Asher *et al.*, 2013; Foraita *et al.*, 2021), as well as for puppies around eight weeks of age (Gazzano *et al.*, 2008; Barnard *et al.*, 2017) and between eight and 10 weeks of age (Bray *et al.*, 2020).

Bray *et al.* (2020) developed a series of tasks for assessing cognition in assistance dog puppies, similar to tasks used by other researchers for working dogs (Lazarowski *et al.*, 2019; Lazarowski *et al.*, 2020; Foraita *et al.*, 2024). Bray *et al.* (2020) noted early emerging communication with humans and that the ability to undertake tasks with visual, olfactory and auditory stimuli was possible in puppies between eight and 10 weeks of age. Bray *et al.* (2021a) reported the heritability of behavioural responses to the social tasks in a population of 375 assistance dogs in the USA. The highest heritabilities were reported for social looking and pointing, both having a heritability of 0.43. Understanding whether the same behaviours were heritable within a guide dog population and whether the same behaviours were associated with guide dog success would advance understanding on the genetic influence for early emerging behaviours, and whether it would be possible to influence future behaviours by breeding.

It is estimated that around 50% of potential assistance dogs do not qualify, largely due to behavioural reasons (Cobb *et al.*, 2015). GDUK have unpublished data showing that approximately 75% of rejections from training and puppy raising were due to behavioural reasons. GDUK have previously trialled behavioural observation assessments at different time points; for puppies between six and eight weeks of age (Asher *et al.*, 2013) for juvenile puppies of five and eight months of age (Harvey *et al.*,

2016) and for dogs in training at approximately 14 months of age (Blythe, 2020). Only one of these assessments has been reviewed for heritability; Puppy Profiling Assessments (PPA; Asher *et al.*, 2013), with heritabilities ranging from 0.02 to 0.25 in over 2,500 puppies (Evans *et al.*, 2015). Since publication, GDUK has not undertaken further heritability analyses on standardised behavioural assessments in puppies or adult dogs. Other guide dog and assistance dog organisations have reported heritabilities for behavioural assessments in young puppies, such as Chen *et al.* (2021a) and Bray *et al.* (2021a). The Bray *et al.* (2021a) heritabilities were limited to human communication tasks and Chen *et al.* (2021a) discussed wider themes of genomic selection for working dogs, so limited results were reported. Therefore, studying heritabilities for a wider group of tasks in a different population will aid in understanding early emerging behaviours in assistance dog puppies. This study aimed to collect practical behavioural data previously associated with cognition and determine whether the behaviours; 1) have a heritable basis and 2) are correlated to the revised Puppy Walker Questionnaire (r-PWQ2).

4.2 Aims and Objectives

4.2.1 Research question

Are data collected through behavioural observations heritable and how does the data compare to data collected through behavioural questionnaires?

4.2.2 Aim

To determine whether: 1) heritable traits can be identified from behavioural observations in puppies as young as seven weeks of age; 2) whether there are correlations to data collected through the r-PWQ2.

4.2.3 Objectives

- Collect behavioural observation data using six previously published tasks in a cohort of guide dog puppies.
- Undertake heritability analysis on the behavioural data.
- Collate 5M r-PWQ2 data for the same subjects and determine if there are correlations between the two data types.

4.3 **Materials and methods**

4.3.1 Study design

A prospective study on a cohort of 399 GDUK puppies between the ages of six and nine weeks was undertaken. Puppies were born between 08/02/2022 and 08/02/2023. Puppies were subjected to six previously published behavioural tasks designed to evaluate puppy cognition (Bray *et al.*, 2020). The tasks were chosen to cover a range of behavioural measures, and all the tasks involved no prior learning or teaching. Behavioural data were collected live during the assessments or coded from videos taken of the assessments using a predefined ethogram. Assessor training was undertaken in December 2021 with the researcher that developed the assessments (Dr Emily Bray) at the Canine Companions' facility in California, USA. Ethical approval was granted by the University of Nottingham Veterinary School's Committee for Animal Research and Ethics (CARE) on 20 December 2021, reference 3501 211130. A full risk assessment for the practical assessments was undertaken by GDUK Health and Safety Team.

4.3.2 Study setting

Behavioural observation assessments took place at the Guide Dogs National Centre (GDNC), Leamington Spa from April 2022 to March 2023. All assessments took place in a small side room with no windows. During

assessments signage and barriers were placed in corridors close to the assessment room to ensure noise levels were kept to a minimum. The assessment room contained a black puppy pen with a gate for entry, known as the testing area (Figure 4.1).

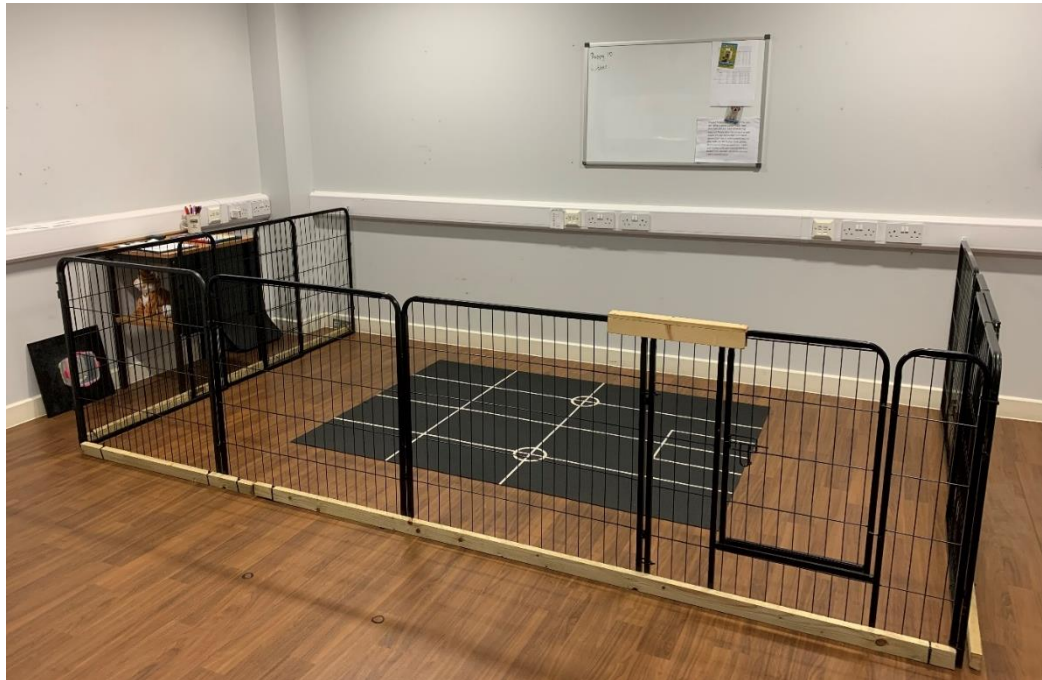


Figure 4.1 Behavioural assessment room. Image shows the puppy pen and mat used during assessments.

A black rubber mat was centred in the testing area with lines painted for standardisation during tasks and to aid with video coding (Figure 4.2).

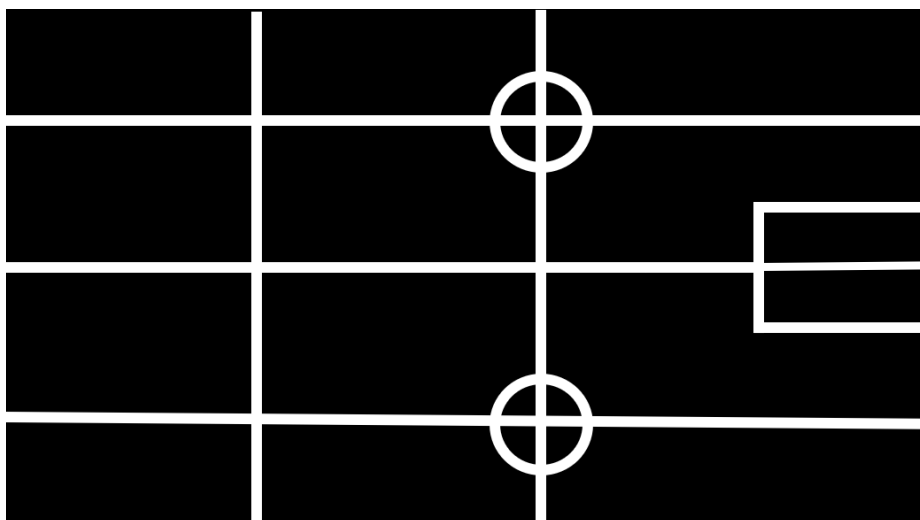


Figure 4.2 Testing mat used for behavioural assessments to show the placement of lines and circles used for equipment, puppy and Experimenter/Handler placement

Two video cameras (Sony 4K FDR-AX53) were permanently housed in the assessment room, one to record a side view (AB angle) and the other to capture the whole testing area (XY angle). The camera for AB was repositioned between position AB1 and AB2 depending on the task and the detail that needed to be captured during the video (Figure 4.3). AB1 angle was used for Unsolvable, Odour Discrimination and Cylinder where a close-up image was required for coding. Angle AB2 was used for Human Interest, Novel Object and Surprising Events where recording the whole pen was required for video analysis. The air conditioning unit was set to 19 degrees for all assessments, which ensured the temperature and conditions were constant for all puppies. In addition, the unit provided a low level of background noise.

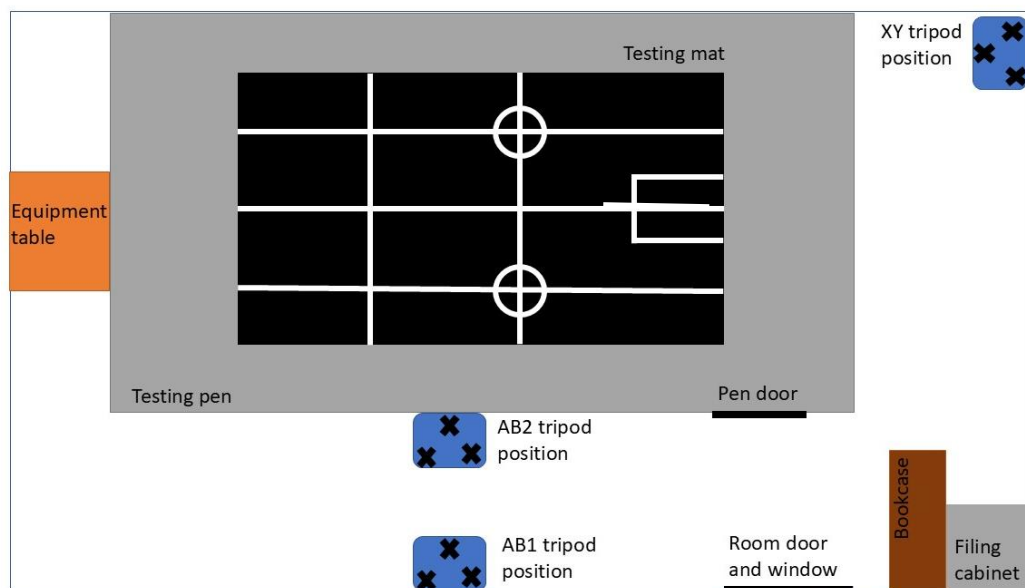


Figure 4.3 Diagram of assessment room setup including pen, mat and camera tripod positioning for AB1 and AB2. Room size was 3.7m x 3.2m, the testing pen was 2.3m x 1.85m and testing mat 1.8m x 1.2m.

The assessments were performed by two individuals, the Experimenter and the Handler. Two Experimenters and three Handlers (including the two Experimenters and an additional staff member within the GDUK

Science Team) collected all the data presented. The Experimenter was responsible for performing the tasks and recording live data (a copy of the data sheets can be found in the Appendix 8.8). The Handler was responsible for timing each trial and letting the puppy go free following a command of 'OK' from the Experimenter. The same Experimenter and Handler was used for each litter of puppies.

4.3.3 Study animals

The animals in this data set were puppies born and raised by GDUK following the same pathway as described in 2.3.2. Puppies were whelped and reared in volunteer homes until around seven weeks of age. At this point they were transported as a litter (without the dam), to the GDNC where they remained in a kennel block for approximately one week. Following this week, puppies were transported around the UK and placed with volunteer puppy raisers.

Puppies were mainly a mixture of Labradors (n=153), Golden Retrievers (n=49), and crosses of these two breeds (n=197; Table 4.1). A total of 399 puppies (206M/193F) were assessed from 36 sires (pups sired ranged from 3 to 26), 69 dams (pups whelped ranged from 1 to 12) and 70 litters (number of puppies per litter ranged from 1 to 10).

Table 4.1 Breed proportions for puppies in behavioural assessments.

Proportion Labrador	Proportion Golden Retriever	Number of puppies	Percentage of total sample
0.000	1.000	49	12.3%
0.063	0.938	25	6.3%
0.125	0.875	27	6.8%
0.250	0.750	10	2.5%
0.500	0.500	86	21.6%
0.563	0.438	16	4.0%
0.625	0.375	23	5.8%
0.750	0.250	10	2.5%
1.000	0.000	153	38.3%
Total		399	100.0%

Litters were selected by breed (Labradors, Golden Retrievers and crosses of these two breeds), with litters marked as 'Potential Breeding Dogs' prioritised to increase the chances of puppies returning as adolescents to the GDNC for breeding dog assessments. A timetable of assessments for the month was planned based on dates litters were due to enter the GDNC. If litters were unwell on the day for planned assessments, assessments were cancelled. Some litters were missed due to puppy illness and a replacement litter used if available.

4.3.4 Assessment protocol

Assessments comprised of six tasks: 1) Unsolvable, 2) Human Interest, 3) Odour Discrimination, 4) Novel Object, 5) Cylinder, 6) Surprising Events. Tasks 1-4 were undertaken in the morning session and Tasks 5-6 in the afternoon session, with the same order of tasks used for all puppies. Puppies were assessed in the same order for AM and PM sessions, so they were given the same rest time between sessions. Morning sessions began around 09:00 and afternoon sessions around 13:00. The term 'task' was used rather than 'test' as there was no pass/fail and no right/wrong response. The order of tasks was determined to ensure puppies received adequate rest between tasks, and that each session ended following the tasks that may provoke a stronger response.

Strict hygiene was followed, adhering to all GDUK cleaning policies including the cleaning of all equipment between sessions and the wearing of scrubs and specific shoes. Before the commencement of each session, the litter was let outside to toilet. Kibble from the daily food allowance was used during the assessments. The 11:00 and 15:30 feeds were taken by the assessors and any food not used during the assessment was given to the puppy at the end of the session. Puppies were identified and carried across the site to the assessment room and were outside for approximately one minute. On the day of assessments, puppies received no additional activities including socialising, to ensure they were not over

stimulated. Each session began with a few minutes to acclimatise the puppy to the room and the Experimenter and Handler. Full assessment methods can be found in the Bray *et al.* (2020) paper. Images of the equipment used can be found in Appendix 8.9. A summary and illustration of each task are shown below.

4.3.4.1 *Unsolvable*

Puppies were familiarised to obtaining a piece of food from a plastic food container attached to a board that did not move. The square box used was a Sistema® food container measuring 16cm x 16cm, which was glued onto a board measuring 35cm x 35cm. Following the familiarisation, four trials were undertaken where the box was locked shut for 30seconds, and the puppy was unable to access the food (Figure 4.4).

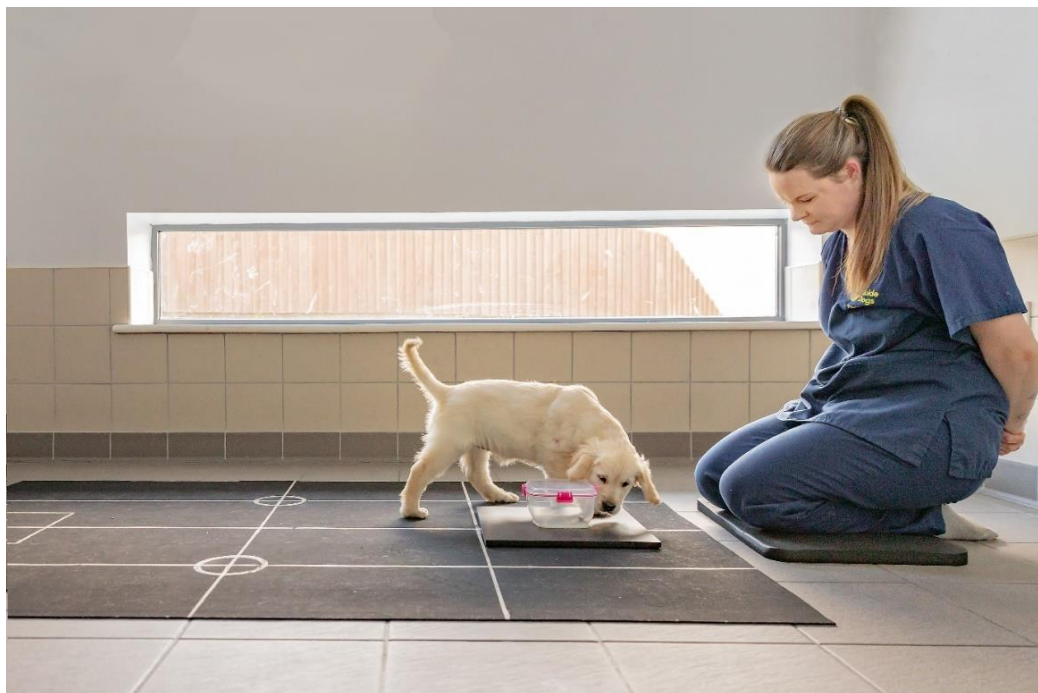


Figure 4.4 Mock up imagery of Unsolvable.

4.3.4.2 Human Interest

Human Interest comprised of two parts. The Experimenter spoke to the puppy from outside the pen, using a set script in a high-pitched, excitable voice for approximately 30seconds. The script was as follows:

“Hi pup! Are you a good puppy? Yes, you are. What a good puppy. Aww, look how cute you are. Look at those big eyes and floppy ears. You’re such a cute puppy! Do you like to play? Are these games fun? You’re coming back later to play with me! We’ll play more games. We’re going to have so much fun. I can’t wait to play with you! Are you the best puppy? Yes, you are. Of course, you are. You’re such a good puppy!”

Following the set script, the Experimenter then entered the testing pen, and the puppy had the choice whether to interact for 30seconds (Figure 4.5).

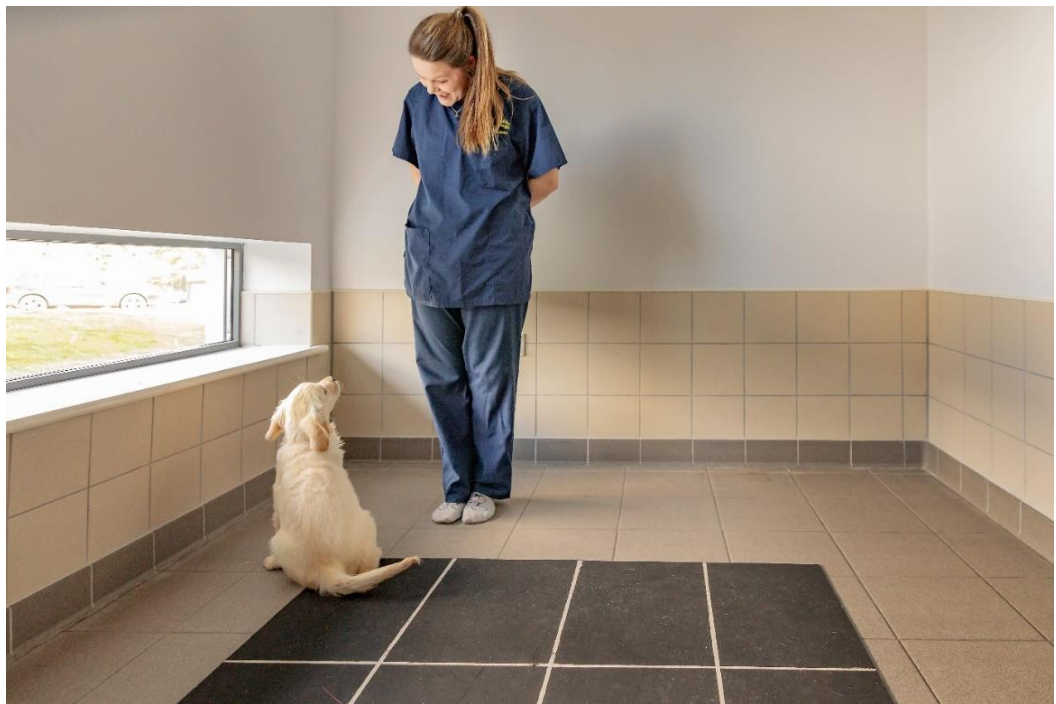


Figure 4.5 Mock up imagery of Human Interest, the Experimenter is talking to a puppy.

4.3.4.3 Odour Discrimination

The task used two L-shaped tubes, referred to as odour elbows (Figure 4.6). Elbow tubes were white corner drainpipes (6cm x 10cm x 13cm) with caps placed on the end consisting of drilled holes to aid in scent diffusion. One elbow was baited with food and the other was left empty, but the two tubes looked identical. To familiarise the puppy to the task, the puppy was shown the elbow with food twice and rewarded with a piece of kibble. Six trials were then undertaken. The two elbows were presented to the puppy to sniff for a few seconds and were then placed in circles on the testing mat. The puppy was released by the Handler and was given 20seconds to explore the elbows. The elbow baited with food was swapped between being in the left or right hand of the Experimenter, with the same order followed for all puppies. Following the 20seconds the puppy was given the kibble, even if the incorrect tube was selected.

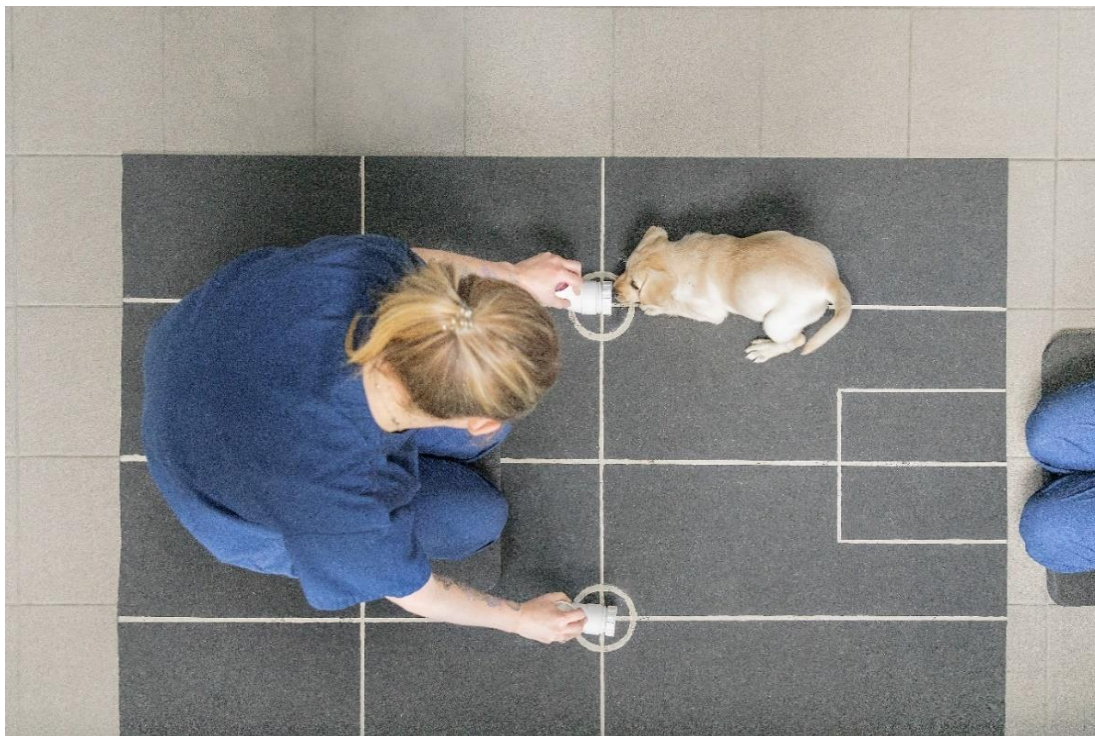


Figure 4.6 Mock up imagery of Odour Discrimination.

4.3.4.4 Novel Object

In Novel Object, an interactive toy cat (SWTOIPIG® orange and white interactive plush cat toy; 30cm x 21cm x 12cm) was placed at one end of the testing mat (Figure 4.7). The Experimenter held the puppy whilst the cat was placed down, and the puppy was then released at the opposite end of the testing mat. The Experimenter left the testing pen and touched the interactive cat as they left which resulted in noises and movement from the toy. Two minutes was timed by the Handler from the moment the puppy was placed down. Both the Experimenter and Handler left the room and watched the puppy through a small glass window in the corridor. Following the two minutes, the Experimenter returned to the room and ensured the puppy was comfortable being next to the cat. The Handler re-entered the room and picked up the puppy without any speaking or interaction with the puppy. The puppy was placed on the floor and called by the Experimenter past the cat. The route taken was recorded using the lines on the mat. The first session ended following this task and the puppy was fed the remainder of their food allowance not used during the assessment. The puppy was then carried back to the kennel block.

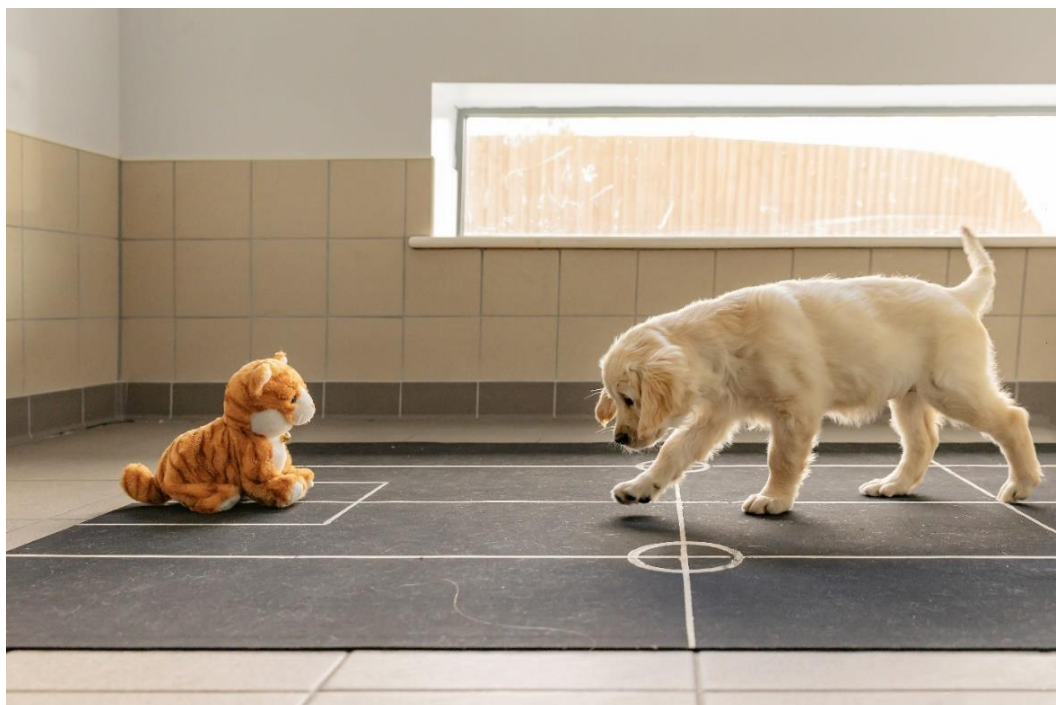


Figure 4.7 Mock up imagery of Novel Object.

4.3.4.5 Cylinder

The Cylinder task used a large cylinder-shaped container attached to a board (Figure 4.8). There were three elements to the task: 1) familiarisation, 2) inhibitory control, 3) detour. First the puppy was shown a small tub (a cardboard cylinder pot measuring 9cm diameter x 3cm height) with a piece of kibble inside. The tub was then used to lead the puppy into an opaque black cylinder twice. The cylinder measured 23cm x 20cm x 20cm and was glued onto a board measuring 50cm x 45cm.

Familiarisation trials then began. The Experimenter showed the puppy the tub with a piece of kibble in, placed the tub in the cylinder and released the puppy. The time taken for the puppy to enter the cylinder, and the path taken to enter the opening (including whether the cylinder was touched) were recorded. To progress from the familiarisation trials, the puppy was required to successfully obtain the kibble from the tub in four out of five trials without touching the cylinder. The puppy was also required to complete familiarisation in 20 trials or in under 10 minutes. Following familiarisation, the puppy completed eight 'inhibitory' and 'detour' trials. For inhibitory, the black cylinder was changed for a clear one and the same measures recorded. For detour, the open side of the cylinder was switched around, so the puppy had to walk to the other side of the cylinder to obtain the kibble.



Figure 4.8 Mock up imagery of Cylinder, showing the detour element of the task.

4.3.4.6 Surprising Events

Surprising Events contained four elements. The first three were exposure to an item that appeared in the testing pen. The items were: 1) a bin bag (Figure 4.9), 2) an umbrella (Figure 4.10), 3) a metal sheet (Figure 4.11). Within elements 1-3, the same methods and measures were used. The final element of the task was to leave the puppy alone for one minute at the end of the trial. The method followed for the three items was the same. The puppy was held by the Handler and the Experimenter stood outside of the pen at the opposite end. The Experimenter then either 1) threw a bin bag full of crumpled paper onto the testing mat, 2) opened an umbrella (1m diameter) or 3) shook a 90cm x 60cm x 0.5cm metal sheet. Items were placed on the testing mat at the opposite end of the pen to the puppy, using grid lines for consistency of placement. Once the item was on the testing mat, 15seconds was timed and whether the puppy approached the item, touched the item or vocalised was recorded. Following 15seconds, the Experimenter entered the pen and walked behind the item. The Experimenter ensured the puppy was comfortable around the item and encouraged the puppy to place their paws on the

item. The Handler picked up the puppy and walked to the opposite end of the testing mat. The puppy was called by the Experimenter and the path taken recorded. Following exposure to the three items, the Experimenter and Handler both left the pen and stood in the corner of the room talking in quiet voices but remained within sight of the puppy. Time spent orientated towards humans and vocalisations were recorded. Following the minute alone, the task ended, and the puppy was given the remaining food from their food allowance. Puppies were then returned to the kennel block.

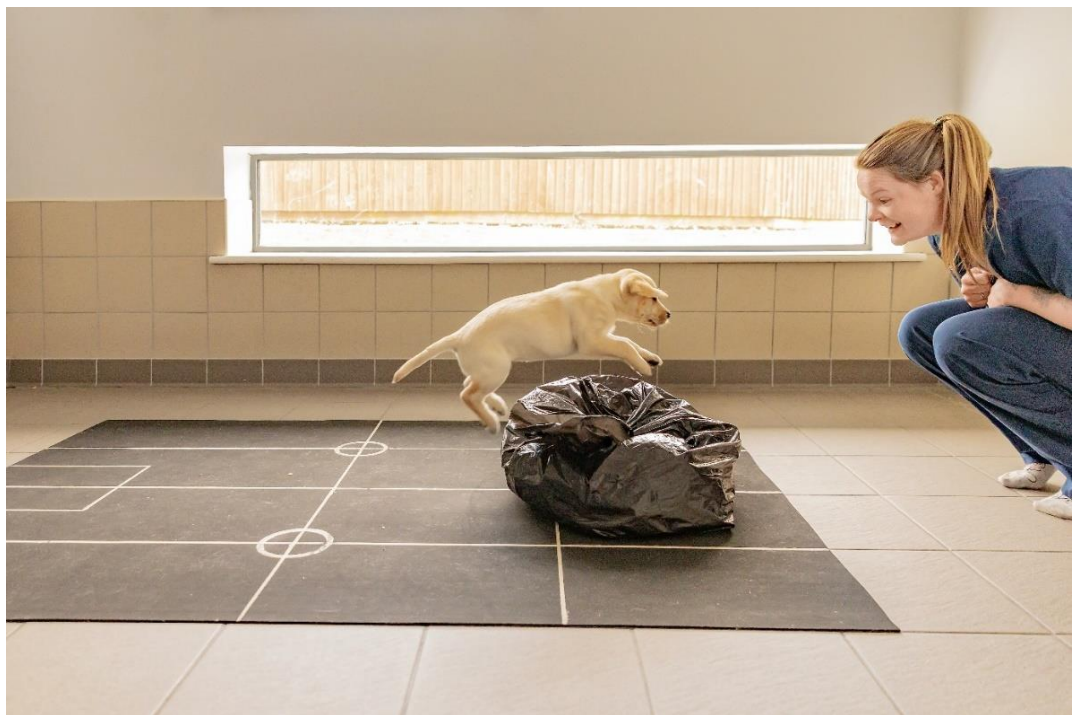


Figure 4.9 Mock up imagery of Surprising Events (bin bag).



Figure 4.10 Mock up imagery of Surprising Events (umbrella).

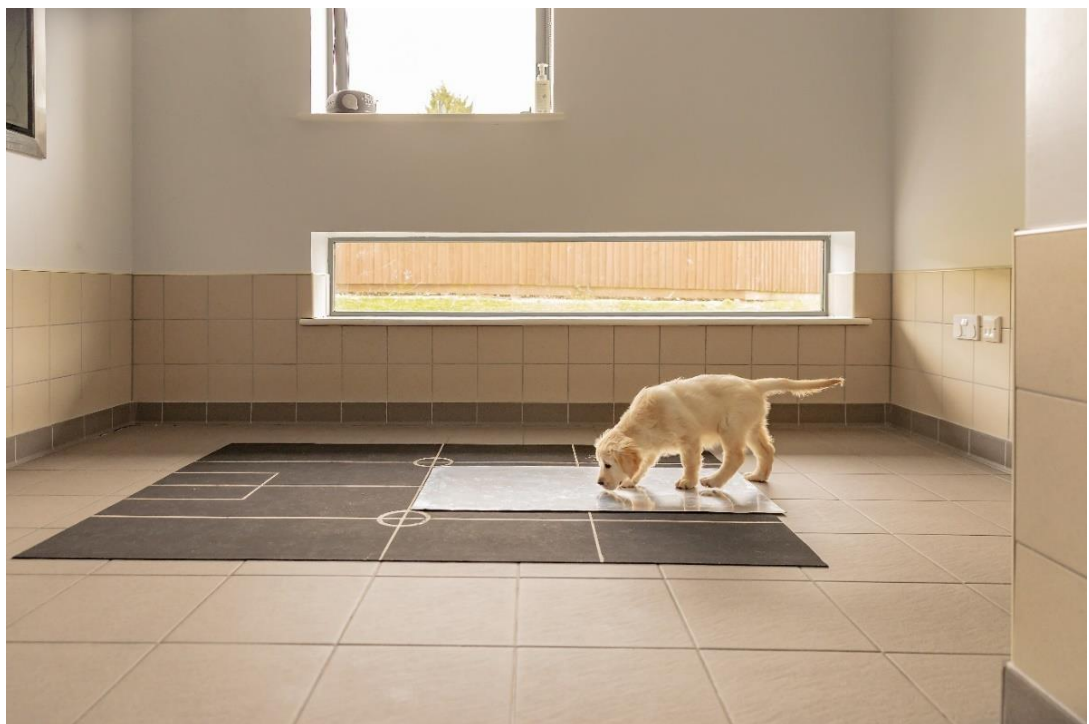


Figure 4.11 Mock up imagery of Surprising Events (metal sheet).

4.3.5 Variables and data sources

4.3.5.1 *Practical assessment data*

Data were collected using a previously validated assessment (Bray *et al.*, 2020) for six individual tasks. Data were either, time, continuous or categorical. Training videos were provided by Dr Bray to demonstrate examples and a pre-existing ethogram/ list of data variables provided to record observations. Measures were recorded using predefined definitions to record the presence of, or intensity of, behaviour. An ethogram was used for Novel Object and Surprising Events. All timings were taken either using a stopwatch or using the timings in the video files. Time variables are reported in seconds, with standard errors. Additional measures of interest were added including placement of front paws on the Experimenter (Unsolvable), orientation towards the Experimenter (Human Interest), presence and intensity of vocalisation (Unsolvable, Human Interest, Odour Discrimination and Cylinder). A summary of the measures collected for each task are shown below.

Unsolvable: The measures recorded live were time the puppy spent making eye contact with the Experimenter (in seconds), vocalisations and whether the puppy placed front paws on the Experimenter. From videos the time spent interacting with the box was recorded using a stopwatch alongside any missed live data.

Human Interest: The measures recorded live were time making eye contact with the Experimenter (in seconds), time interacting with the Experimenter (in seconds) and vocalisations. From videos the time spent orientated towards the Experimenter when being spoken to were recorded using a stopwatch alongside any missed live data.

Odour Discrimination: The measures recorded live were the first and last elbow choice and whether the puppy switched between the elbows. This resulted in a score for whether the puppy chose the correct elbow first and last. Any repeats and vocalisations were also noted. From videos, the time

spent interacting with each elbow was recorded using a stopwatch alongside any missed live data.

Novel Object: All measures for Novel Object were coded from videos using a predefined ethogram (see Table 4.2). Some measures were recorded as 'Y' when the puppy showed the behaviour and 'N' when the puppy did not show the behaviour. The measures recorded were: time spent approaching the cat, time spent in contact with the cat, time spent orientated towards the cat, grab (Y/N), crouch (Y/N), play bow (Y/N), pant (Y/N), time of first vocalisation, latency to first vocalisation, time spent vocalising, whether the puppy was vocal towards the cat or the door, and the type of vocalisation (groan, whine, yelp/bark, howl), which provided a vocal intensity score. When the Experimenter re-entered the pen, the latency to approach, an approach score (see ethogram) and whether a lure was needed for the puppy to approach the cat were recorded. For the recall, the time of release and time to approach the Experimenter were recorded to produce an approach latency time in seconds. The path to approach and if a lure was required were also recorded.

Table 4.2 Novel Object ethogram, 'E' indicates Experimenter and 'H' indicates Handler

Timing	
Alone Start	When E places dog onto testing mat
Alone End	2 min after Alone Start
Behaviour	
Grabs	Dog bites/mouths cat (Y/N)
Crouch	Crouch/ducking (downward movement of body/head to crouched position or walking backwards away from cat to sitting) (Y/N)
Play bow	Front legs out in front and chest low to the ground as if about to lie down, but rear end remains up (Y/N)
Pant	Dog open mouth pants during the session (Y/N)
Approach Cat	Time that puppy approaches cat, defined as all 4 paws on the mat past the 2nd horizontal white line in front of the cat, or in the analogous area behind the cat - i.e. all 4 paws in the red square drawn on the screenshot to the right
Leave Cat	Time that puppy leaves the cat, defined as when all 4 paws are no longer in the red square
Contact	Amount of time that pup spends sniffing and/or with any part of its body in contact with the cat
Lat To Contact	Amount of time until dog first contacts the cat. If dog never contacts the cat, it receives the highest score of 120s
Orient	Amount of time that pup spends with face oriented toward the cat
Vocalisation	
Latency to Vocalisation	Amount of time until pup makes first sound of the type listed below. If dog never vocalizes, it receives the highest score of 120s
Vocalisation Intensity	Intensity of vocalization: Groan/growl, whine, yelp/bark, howl
Cat Vocal	Pup vocalizes while orienting at cat (Y/N)
Door Vocal	Pup vocalizes while orienting toward the back of the pen (i.e. the direction where E left from / the right side of the screen when viewing the 'AB' camera angle) (Y/N)
Vocal Intensity	
Groan/growl	An audible low-intensity noise, ranging from soft grunts to louder groans and moans, mouth closed and/or play growls. (Y/N)
Whine	Elongated high-pitched vocalizations, mouth closed. (Y/N)
Yelp/Bark	Sharp, staccato high-pitched vocalizations, mouth open. (Y/N)

Howl	Elongated high-pitched vocalizations, mouth open, head back. (Y/N)
WHEN E RE-ENTERS (following 2 min 'alone' period) - "2nd approach" period	
Latency	Time (s) from when E is standing by object to when puppy approaches object. If the puppy is already at object or follows E to object immediately, the latency is 0
Score	1 - puppy is at object when E enters or follows E to object
	2 - puppy comes with no food
	3 - puppy comes with food, but no luring
	4 - puppy comes when lured (i.e. E has to leave cat to show puppy food)
	5 - puppy never approaches the cat
Luring	E leaves cat to show puppy food in an effort to get puppy to reapproach (Y/N)
WHEN H RE-ENTERS (following E re-entry) - "3rd approach" period	
Latency	Time (s) from when H releases puppy to when puppy approaches object (using same approach definition as during alone period).
Pathway	1 - puppy's front paws stay in the blue to green zone for the entire reapproach (middle of mat straight over cat to E)
	2 - puppy's front paws stay in the blue to yellow zone for the entire reapproach (middle of mat and then around cat to E)
	3 - puppy's front paws go into the orange zone for part of reapproach, but never leave mat
	4 - puppy's front paws go into the red zone for part of reapproach (leaves mat)
Luring	E leaves cat to show puppy food in an effort to get puppy to reapproach (Y/N)

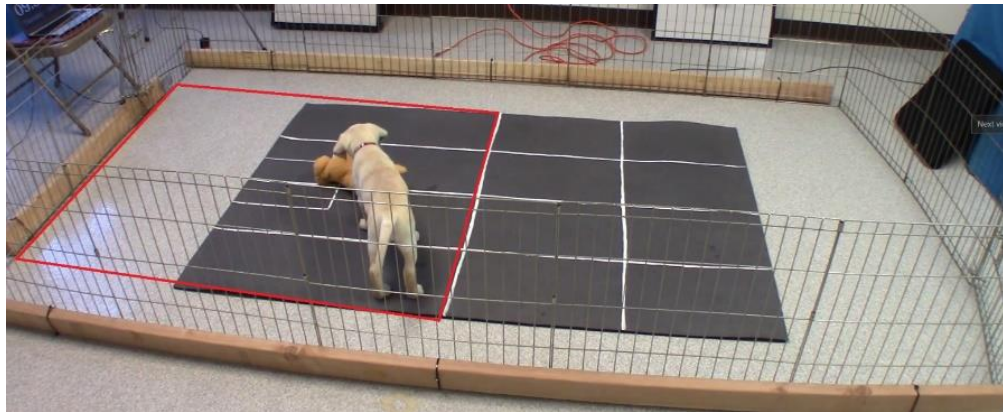


Figure 4.12 Red square to define Novel Object proximity to cat

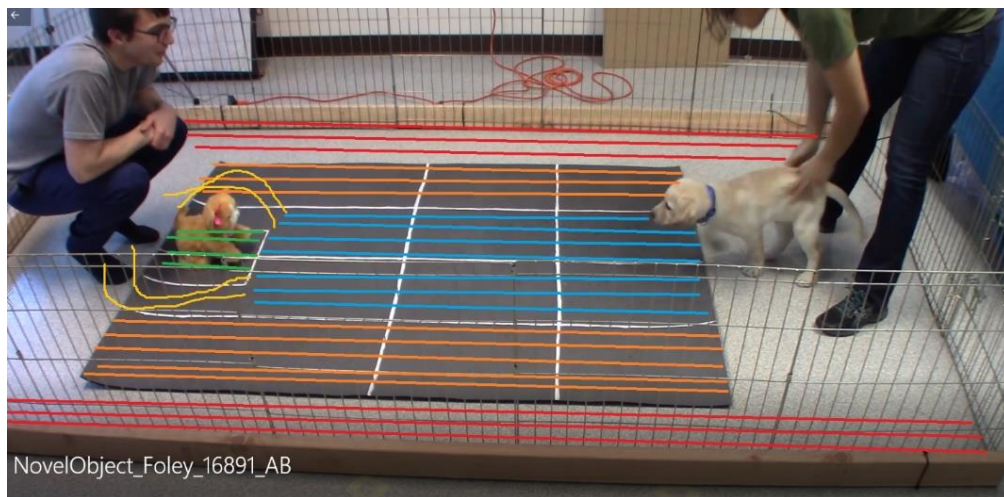


Figure 4.13 Scoring protocol for Novel Object reapproach zones

Cylinder: All measures for Cylinder were collected live unless data were missed, or when the Experimenter and Handler were unsure of a touch. In these cases, videos were used to code the data. For familiarisation trials, the data recorded were whether the puppy correctly entered the cylinder without touching (1/0), what part of the cylinder the puppy touched (front, top, side or back), and the latency to solve. For inhibitory and detour trials, the same measures were recorded as for the familiarisation trials, in addition to whether the approach was direct (defined as the puppy going straight into the cylinder on the open side) and whether there were any vocalisations.

Surprising Events: All data were coded from videos using a predefined ethogram (see Table 4.3). Data were collected for three separate objects and then for one minute where the puppy was left alone. For the three objects (bin bag, umbrella, and metal sheet), the same data were collected: the initial response to the object, the time the puppy approached the object, the approach score, path to approach and the minimum distance approached in 45 seconds. The time in contact with the object and the vocalisations were also recorded. On the reapproach, the time to recall, the path the puppy took and the reapproach minimum distance were recorded. For the minute alone, the time orientated towards the Experimenter/Handler the time of first vocalisation and the intensity of vocalisations were recorded.

Table 4.3 Surprising Events ethogram

Time Stimulus	Time each trial starts, defined as when stimulus is placed on ground .	
Response	<i>The initial reaction score of the dog to the stimuli. Beginning within one second - immediate response. The most extreme reaction visible until stimulus hits the ground.</i>	
1	No detectable reaction (attention, orienting (turning head, ears perking) is fine. dog may sit if no flinch or startle)	
2	Flinch/startle (without lowering of body, but some movement is ok- including a small step back)	
3	Crouch/duck (downward movement of body/head) without major displacement and maintaining general body orientation	
4	Rapid avoidance response away from stimulus (can be paired with a crouch and/or a change in general body orientation)	
NOTES:	<i>If the dog's only response is to move forward, they get a 1. If they crouch and move forward, they get a 3.</i>	
Approach measures - to be used for each stimulus		
Approach score		
1	Approach 0-15	If the dog <i>initially</i> approaches the stimulus (steps all four paws into region bounded in orange denoted in the above photo OR touches with nose) before 15 seconds have passed, the dog receives a 1. This means the dog will have approached with no encouragement from the experimenter.
2	Approach 15-30	If the dog <i>initially</i> approaches the stimulus after 15 seconds have passed but before 30 seconds have passed, the dog receives a 2. This means the dog will only have approached after receiving verbal or food lure encouragement from the experimenter.
3	Approach 30-45	If the dog <i>initially</i> approaches the stimulus after 30 seconds have passed, the dog receives a 3. This means the dog will only have approached after extensive verbal encouragement and food lure from the experimenter.
4	Approach Never	If the dog is never convinced to approach the stimulus during the 45 second trial, the dog receives a 4. This means the dog was never convinced to approach the stimulus, even after verbal and food coaxing.
Approach path	From release to first approach (if first approach within 15 seconds)	
1	dog's front paws stay in the blue, green, or yellow zone for the entire approach (middle of mat straight to/over/around stimulus)	
2	dog's front paws go into the orange zone for part of approach, but never leave mat	
3	dog's front paws go into the red zone for part of approach (leaves mat)	
4	IF NOT WITHIN THE FIRST 15 SECONDS	
Notes:	**WHOLE PAW NEEDS TO LEAVE THE ZONE AND ENTER ANOTHER ZONE TO COUNT* *If object is not centred on the edge of the mat, shift orange approach box mentally	
Time to first contact (sec)	The first time that the dog makes actual contact with the stimulus (any time prior to being picked up for reapproach) *If they sniff close enough that no space is visible between nose and object, count as contact	

Amount contact (sec)	Amount of time dog is in contact with the stimulus in the approach phase only (within first 15 secs, maximum 15 secs)	
Approach Minimum distance from stimulus	CODED FOR ANY TIME DURING THE FIRST 45 Seconds of Approach Phase	
1	On top of stimulus with weight on paws with all 4 paws (at the same time)	
2	On top of stimulus with two paws, with weight on both - for bin bag minimum distance, both paws must be two paw widths from the edge of the bag (toward the centre of the bin bag) to be a 2.	
3	Touching stimulus with any body part but not standing on to - for bin bag minimum distance, if paws are just on the edge of the bag, it's a 3.	
4	Not touching but within one rectangle	
5	Not touching and not within one rectangle	
Vocalisation		
Groan/growl	An audible low-intensity noise, ranging from soft grunts to louder groans and moans, mouth closed and/or play growls (Y/N)	
Whine	Elongated high-pitched vocalizations, mouth closed (Y/N).	
Yelp/Bark	Sharp, staccato high-pitched vocalizations, mouth open (Y/N).	
Howl	Elongated high-pitched vocalizations, mouth open, head back (Y/N).	
Reapproach path	From release to first contact with experimenter	
1	dog's front paws stay in the blue to green zone for the entire reapproach (middle of mat straight over stimulus to E)	
2	dog's front paws stay in the blue to yellow zone for the entire reapproach (middle of mat and then around stimulus to E)	
3	dog's front paws go into the orange zone for part of reapproach, but never leave mat	
4	dog's front paws go into the red zone for part of reapproach (leaves mat)	
Reapproach Minimum distance from stimulus	FOR THE FIRST 45 SECONDS OF REAPPROACH (FROM WHEN RELEASED)	
1	On top of stimulus with weight on paws with all 4 paws (at the same time)	
2	On top of stimulus with two paws, with weight on both *For bin bag minimum distance, both paws must be two paw widths from the edge of the bag (toward the centre of the bin bag) to be a 2.	
3	Touching stimulus with any body part but not standing on top *For bin bag minimum distance, if paws are just on the edge of the bag, it's a 3	

4	Not touching but within one rectangle
5	Not touching and not within one rectangle
Alone time (1 min at end)	
Time alone start	Once both E and H have exited the arena with both feet
Orientation	
Orientation Start	Start each time the dog faces/orients toward the E - Must be facing E - if head turned away, do not count
Orientation End	End each time the dog faces/orients away from the E
Vocalisation	
Time first vocalisation	Mark the time when the dog first vocalises
Time alone end	Defined as one minute after the time alone start (E may again address the dog after this point)

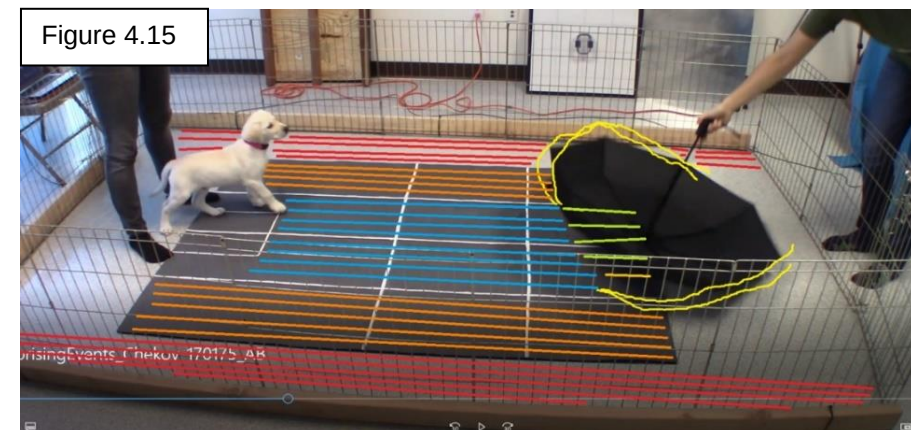
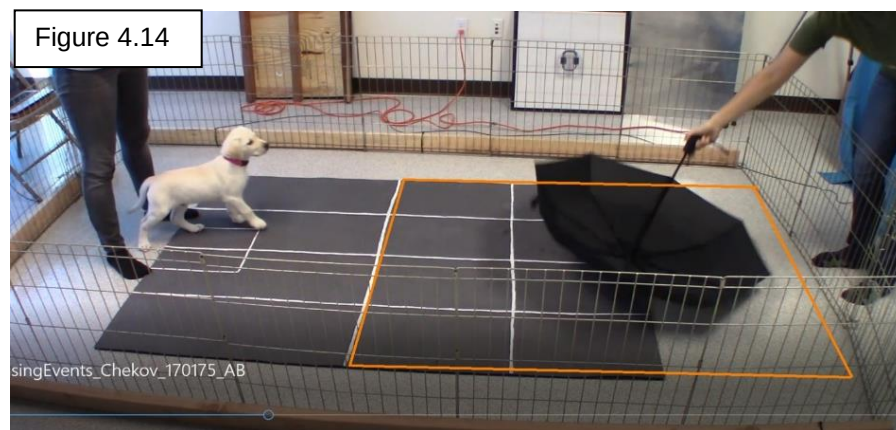


Figure 4.14 Red square to define Surprising Events proximity to object
Figure 4.15 Scoring protocol for Surprising Events reapproach zones

The six tasks covered a range of behavioural responses, including aspects of cognition and problem solving. The results were grouped based by behavioural measure types rather than by task. The groups of behaviour include 1) human engagement, 2) problem solving, 3) interaction with a novel item, 4) surprise response and recovery and 5) vocalisations.

Data were also collected from GDUK's electronic database including puppy name and ID, date of birth, breed, sire ID, dam ID and litter ID.

4.3.5.2 Pedigree data

An existing GDUK pedigree, provided and coded by Dr Tom Lewis, was used for heritability analyses. Within the pedigree file, dogs were assigned a unique id (*pid*). The *pid* for the individual dog, the sire and dam were collated in one data file for analyses. The final pedigree file contained 64,270 individuals born between 1901 and 2023, of which 33,470 were female and 30,800 were male. The most common breeds represented in the pedigree were Labradors (n=24,994), Golden Retriever x Labrador or Labrador x Golden Retrievers (n=15,697) Golden Retrievers (n=8775) and German Shepherd Dogs (5090). There were 5579 unique sires and 9618 unique dams, with unknown sire information for 6173 dogs and unknown dam information for 6954 dogs. The mean number of offspring per sire was 10.4, with a minimum of 1 and a maximum of 622. The mean number of offspring per dam was 6.0, with a minimum of 1 and a maximum of 61.

4.3.5.3 5M r-PWQ2 data

Complete r-PWQ2 responses were collected for 308 puppies (162M/146F) at 5M and were extracted from the GDUK online database (GDIR). 5M r-PWQ2 scores followed the same component groupings and calculation methods reported in Chapter 2.

4.3.6 Bias and study size

Study size was limited by the number of puppies available at the GDNC. Golden Retrievers, Labradors and crosses of these two breeds were prioritised to ensure a sufficient sample could be collected for analyses. Data were collected for an additional 14 Golden Retriever x German Shepherd Dog puppies; however, these data were excluded from further analyses due to the different breed mix and small sample size for the breed.

It was possible to undertake assessments on a maximum of six puppies per day, therefore not every puppy within a litter were included. Puppies were chosen at random from the litter whilst ensuring a representative mix of male and female puppies from each litter. The same order for puppies was used throughout the day so puppies had the same rest period between AM and PM sessions. This resulted in some puppies being fed closer to the first session than others, but as food from two meals was used during the assessments, food was given at the same intervals between sessions. There is the potential for the timing of the first feed (7:30am) to have impacted interaction with food based tasks for some puppies which could not be controlled for.

Puppies were not always assessed on the same day post entry to GDNC, as the day of assessment was determined by many factors including assessor availability, visitors to the site, puppy health and the number of puppies in the kennel block available for assessment in a given week. Assessments were also missed due to staff holidays or extremes of weather (a heatwave of 37°C and cold temperatures of -10°C were experienced during assessments). Extremes of weather affected assessments as puppies had to be carried outside for a short walk to get from the block to the assessment room. To ensure puppy welfare was always maximised, the assessments were cancelled when Dog Health and Welfare experts at GDUK stated they should not go ahead.

Variation between assessors was minimised by using a small number of individuals (n=2) to code behavioural observations from videos, similarly the same two individuals were trained as Experimenters to furthermore reduce variations in assessment methods. The analysis of tasks were split between individuals to reduce differences in coding between the tasks. A scoring protocol was provided by Dr Bray (the researcher that developed the assessments) and an ethogram was used to ensure consistency in coding. Inter and intra-rater reliability analyses (Chapter 5) were undertaken to determine coder differences, and following these results, several tasks were recoded.

4.3.7 Quantitative variables and statistical methods

4.3.7.1 *Heritability*

Data from all tasks were collated and 'NA' was used if a puppy did not undertake a specific task or when data were missing. All tasks with a time were converted to seconds using the following code in R:

```
x=c(i)
for (i in x){ s=as.numeric(substr(data[,i],4,5))
  m=as.numeric(substr(data[,i],1,2))
  data[,i]=(m*60)+s }
```

The pedigree file was loaded into R and the pid (unique id within the pedigree) and proportion of Labrador were extracted as a numeric value from 0 to 1. Where 0 indicates the puppy has no Labrador within their breed (i.e., the puppy is a pure Golden Retriever) and 1 indicates a pure Labrador.

A heritability model was run for all individual behaviours (n=96), using the fixed factors of sex, proportion of Labrador and age in weeks. Random factors included were the unique id within the pedigree and inverse of the pedigree. The model was set up to omit missing values (NA values) and to have a maximum iteration of 50 for convergence of the log-likelihood.

The following model was run in R using the *asreml* () function for each behavioural response:

$$y = 1\mu + X\beta + Z\alpha + \varepsilon$$

where y is a behaviour trait, μ is the overall mean, β is a vector of fixed effects with X as the corresponding incidence matrix, Z is the incidence matrix for the random additive polygenic effect, α is a vector of random additive polygenic effects distributed as for the pedigree-based estimates and ε is a vector of residual errors. Fixed effects included in all models were the sex and breed (proportion of Labrador) of the age of the puppy in weeks.

Heritability analyses were undertaken using the ASReml package (version 4.2) within R (version 4.3.1). The significance of fixed effects (breed, sex, age) was determined by dividing the Z ratio variance by the standard error. Values greater than 1.96 were considered significant.

A histogram was created for all behavioural measures from the assessment (96 measures) to determine distribution and variation in the data set. Heritability models were run for all 96 variables. All measures were reviewed for variation and where both the average time and total time were recorded, only the average time was used in analyses due to replication present between the total and average time. For variables with less than 10% variation in the sample, and for variables with a large proportion of missing data, these measures were not included.

4.3.7.2 *Principal component analysis*

It was not possible to run a principal component analysis (PCA) with all the behaviour variables measured due to the small size of the data set. A PCA was run for the top 22 heritable behaviour measures. A Varimax rotation was applied and component values above 0.4 were retained. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of

Sphericity were used to determine whether a PCA was appropriate for use in each data set. All statistics for PCAs were undertaken using SPSS version 28 (SPSS Inc., Chicago, IL).

4.3.7.3 Correlations to 5-month r-PWQ2

Five-month r-PWQ2 data was available for 308 of the 399 puppies in the sample. The r-PWQ2 behaviour components were calculated following the same methods reported in Chapter 2 (see section 2.3.6.2). An average of the response to each question in each behaviour component grouping was used for analyses. The minimum, maximum, mean and standard error of the mean were calculated for each of the 10 components.

A Kendall's tau-b correlation was run to compare the top 22 heritable behaviour measures to the 10 r-PWQ2 components. Following the PCA, extracted behaviour components were then compared to the 10 r-PWQ2 components also using Kendall's tau-b. Correlation comparisons were undertaken using SPSS version 28 (SPSS Inc., Chicago, IL).

4.4 Results

4.4.1 Assessment data

Age at assessment ranged from 6.3 to 10 weeks of age, with a mean age of 7.4 weeks at assessment (± 0.04). Out of the 399 puppies, a total of 288 (72.2%) completed all six tasks. The time for assessment for all puppies ranged from 17 to 69 minutes, with a mean time of 43.2 minutes (± 0.4). For puppies that completed all tasks, the mean time for completion was 43.9 minutes (± 0.3), which was spread over two sessions (an AM and a PM session). Experimenter 1 assessed 305 puppies (76.4% of puppies), and Experimenter 2 assessed 94 puppies (23.6% of puppies).

During the year of data collection, an average of 33.3 ($\pm$ S.D.14.9) assessments were undertaken per month (Figure 4.12).

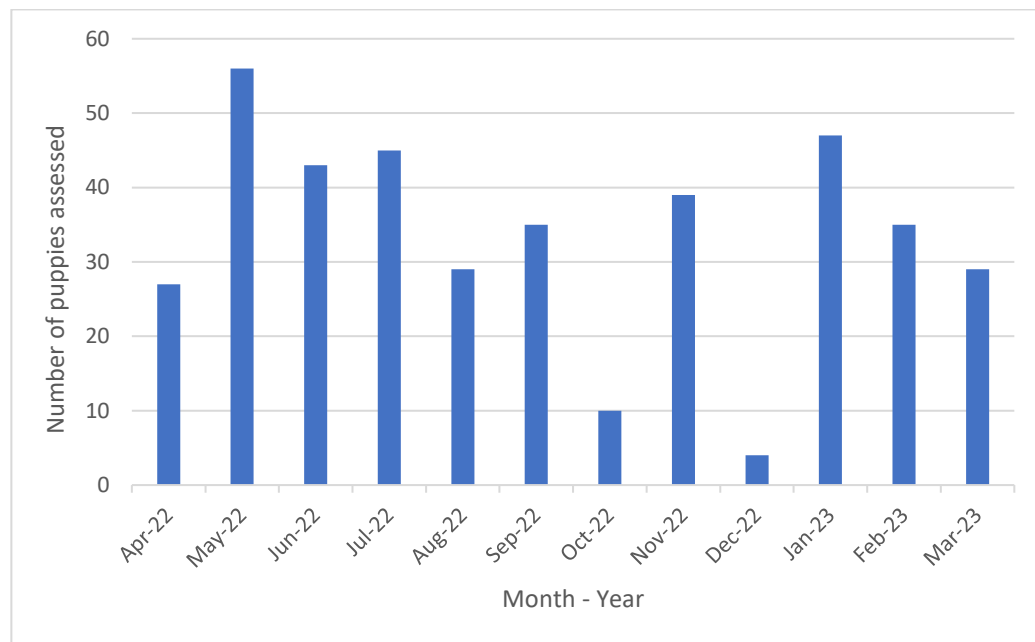


Figure 4.16 Puppy assessments undertaken by month from April 2022 to April 2023.

4.4.1.1 Task completion

Data for the Human Interest task were collected for all 399 puppies (100.0%), and data for the Novel Object task were collected for 398 puppies (99.75). Data for the Unsolvable task were collected for 383 puppies (96.0%), as 16 puppies did not progress past the familiarisation stage. Complete data for the Odour Discrimination task were collected for 385 puppies (96.5%); six puppies did not undertake the task, and eight puppies did not complete all six trials (data for complete trials was included in analyses). Data for the completed Cylinder task were collected for 287 puppies (71.9%) with 356 puppies completing just the familiarisation section of the task and 293 puppies completing both the familiarisation and inhibitory sections. Complete data for the Surprising Events task were collected for 367 puppies (92.0%). Surprising Events data were missed due to puppy illness, or the puppy having left the centre, and therefore no afternoon session took place. Three puppies did not fully

complete the Surprising Events as the task was aborted due to either a fear response shown by the puppy or noise distractions occurring outside of the assessment room that influenced the puppy's behaviour.

4.4.2 Behavioural analysis

4.4.2.1 *Human engagement*

Mean looking time across the four trials of Unsolvable was 0.7 seconds (± 0.1), with a range of 0.0 to 7.3 seconds. Approximately half of puppies (49.1%; 188 puppies) did not place two front paws on the Experimenter during any of the four trials. Puppies on average spent less than one trial with their paws on the assessor (mean 0.9 ± 0.1).

Three trials of 30 seconds were completed for Human Interest. The mean time spent making eye contact with the Experimenter ranged from 0.0 to 19.5 seconds with a mean of 3.3 seconds (± 0.2). Mean orientation towards the Experimenter ranged from 1.6 to 26.7 seconds, with a mean of 13.1 seconds (± 0.3). Following being spoken to, the mean time spent interacting with the Experimenter ranged from 0.0 to 30.0 seconds, with a mean of 25.2 seconds (± 0.3). During the minute of alone time in Surprising Events, the time spent orientated towards the Experimenter and Handler was 20.4 seconds (± 0.7) with a range of 0.0 to 58.0 seconds.

4.4.2.2 *Problem solving*

The mean manipulation time across the four trials for Unsolvable was 11.2 seconds (± 0.2) with a range from 0.0 seconds to 24.1 seconds. For Odour Discrimination, the correct elbow was selected by puppies as their first choice for or 55% of the population. Choosing the correct elbow at the end of the trial was found for 79% of puppies. Time spent at the correct elbow over the six trials ranges from 3.9 to 116.4 seconds, with a mean of 72.7 seconds (± 1.1).

For puppies that completed the task (356 completed familiarisation, 293 completed inhibitory control and 287 completed detour), the mean time to solve the Cylinder task was 6.8 seconds (± 0.3) for familiarisation, 4.6 seconds (± 0.2) for inhibitory control and 8.3 seconds (± 0.2) for detour. The number of warmups required to move from the familiarisation task to inhibitory was a mean of 8.5 trials (± 0.2), and the range was between four and 20. On average (mean), puppies correctly entered the cylinder 39% of the time (0.4 ± 0.0) and detour 18% of the time (0.2 ± 0.0).

4.4.2.3 Interaction with novel item

The mean time taken for puppies to first approach the cat was 8.8 seconds (± 1.0), with a range of 1.0 to 120.0 seconds, and the number of approaches ranged from 0 to 10 with an average of 3.3 approaches (± 0.0) per puppy. The total time in close proximity to the cat was a mean of 53.9 seconds (± 1.5) and time spent orientated towards the cat was a mean of 60.3 seconds (± 1.0). Mean latency to contact was 14.7 seconds (± 1.5) with a mean of 23.1 seconds (± 1.0) spent in contact with the cat. There were 38.7% ($n=154$) of puppies that grabbed the cat, 49.5% ($n=197$) that crouched towards the cat and 15.3% ($n=61$) of puppies displayed a play bow.

During the approach when the Handler re-entered the room and following a recall, only one puppy required a lure. Puppies were commonly present at the cat by the time the Experimenter had her feet behind the item (57.3%, $n=228$). Mean time to approach the cat with the Experimenter present was 1.9 seconds (± 0.2) and following a recall was 3.8 seconds (± 0.2). The most frequent path taken to approach the cat following a recall was 2 (dog takes a direct approach to the cat but does not place front paws directly next to the cat), which was the case for 205 puppies (51.5%).

4.4.2.4 Surprise response and recovery

The mean initial surprise response to the item was 1.4 (± 0.0) for the bag, 2.9 (± 0.0) for the umbrella and 3.0 (± 0.1) for the sheet, where a higher score represents a higher startle response. Mean latency to approach the item was highest for the bag (15.1 ± 0.7 seconds), followed by the sheet (14.1 ± 0.6 seconds) and lastly the umbrella (9.2 ± 0.6 seconds). Mean latency to contact the item was lowest for the umbrella (6.3 ± 0.5 seconds), followed by the bag (13.5 ± 0.7 seconds) and lastly the sheet (13.9 ± 0.6 seconds). Total time contacting the item showed the same trend, the highest contact was made with the umbrella (6.2 ± 0.2 seconds), then the bag (5.5 ± 0.3 seconds) and lastly the sheet (1.4 ± 0.1 seconds).

The path taken to approach each item was recorded for the first and second approach, with a higher score indicating that a greater width was taken to avoid the stimuli. The mean path score was lowest for the umbrella (2.0 ± 0.1), followed by the bag (2.4 ± 0.1) and then the sheet (3.1 ± 0.1). For the reapproach the lowest was for the sheet (1.9 ± 0.1), then the umbrella (2.0 ± 0.1) and lastly the bag (2.1 ± 0.0). The minimum distance score (where a smaller score indicated the puppy is closer and in greater contact with the item) was lowest for the umbrella (1.1 ± 0.0), then the sheet (1.2 ± 0.0) and lastly the bag (1.7 ± 0.1). Scores for the second reapproach minimum distance were similar between the three elements (mean score bag 1.1 ± 0.0 for the sheet – 1.1 ± 0.0 and for the umbrella 1.1 ± 0.0). Mean latency to reapproach the item following a startle was lowest for the sheet (3.6 ± 0.1 seconds) and highest for the bag (4.5 ± 0.3 seconds).

4.4.2.5 Vocalisations

Puppies commonly did not vocalise in Odour Discrimination (91%, $n=358$), or Cylinder (96.6% ($n=282$) in Inhibitory and 93.0% ($n=267$) in Detour; vocalisations in familiarisation were not recorded). Most puppies did not vocalise during the Unsolvable task (67.1%; $n=257$). The most frequent vocalisation during the Unsolvable task was a whine, observed in 75

puppies (19.6%). During Human Interest, 19.3% of puppies (n=77) showed no vocalisations during the talking part of the task, and 70.2% (n=280) showed no vocalisations during the interaction part of the task. The most common vocalisation during Human Interest talking was a whine (41.1%, n=164), followed by a bark/yelp (29.6%, n=118). The most common vocalisation during the interaction section was also a whine (17.3%, n=69).

During Novel Object, 85.7% (n=341) of puppies vocalised towards the cat and 69.1% (n=275) vocalised towards the door (where the Experimenter and Handler had exited the room). Of the 398 puppies that completed Novel Object, 47.7% groaned (n=190), 88.7% whined (n=353), 33.4% barked or yelped (n=133) and 20.1% howled (n=80). For Surprising Events, the vocalisations during each of the four elements of the task were recorded. For the bag appearance and the opening of the umbrella, 58 puppies vocalised (15.7%), with the most common vocalisation being a whine (n=28 (7.6%) for the bag and n=44 (11.9%) for the sheet). More puppies vocalised in the sheet element of Surprising Events when compared to the bag and umbrella, with 54.9% (n=202) vocalising, and the most common being a whine (n=170, 46.2%). During the minute alone in Surprising Events, the mean latency to vocalisation was 42.7seconds (± 1.01), with 59.7% (n=219) vocalising during the task and whining being the most frequent vocalisation (n=161, 43.9%).

4.4.3 Breed, sex and age effects

Due to the difference in age at assessment and the presence of varying breeds and sexes in the sample, analysis was undertaken to understand whether these factors had any influence on the behavioural responses displayed. A Z-ratio of greater than 1.96 was considered significant.

Breed effects were present for umbrella reapproach score (Z-ratio -3.07; highest scores for 0.75 Lab and lowest for 0.25 Lab), umbrella latency to

contact (Z-ratio -2.65; highest latency for 0.25 Lab and lowest for 0.75 Lab), sheet reapproach path (Z-ratio -2.60; highest scores for 0.75 Lab and lowest for 0.25 Lab) and umbrella reapproach path (Z-ratio -2.33; closest reapproach for Lab and furthest for 0.25 Lab) in the Surprising Events task. Breed also influenced whether the puppy groaned during the Novel Object (Z-ratio 3.82). Sex effects were present for the maximum vocalisation category in Human Interest (Z-ratio 2.71) and Odour Discrimination (Z-ratio 2.32), with males having a higher vocalisation category for both. In Surprising Events, bag reapproach path (Z-ratio 2.19; males had a winder reapproach), contact time with the umbrella (Z-ratio 1.99; males had longer contact) and sheet latency to reapproach (Z-ratio 1.98; males took longer to reapproach) showed sex effects,

Age had the largest overall effect on behavioural measures. Significant effects were present for total and average looking time in Human Interest (Z-ratio 5.02; older puppies made eye contact for longer), and the number of vocalisations in the Unsolvable task (Z-ratio -2.00; younger puppies vocalised more). The last choice being correct in Odour Discrimination (Z-ratio 1.97; higher for older puppies) and time spent at the incorrect elbow (Z-ratio -2.21; longer for younger puppies) were influenced by age. The tasks most impacted by age were Cylinder and Surprising Events.

For Cylinder, inhibitory score (Z-ratio 2.92), detour score (Z-ratio 2.17) and the latency for all three sections (familiarisation average latency Z score -5.02, inhibitory average latency Z score -2.12 and detour average latency Z score -4.00) were affected by age. Older puppies were quicker and scored higher than younger puppies. Similarly, the number of warm-ups required to move onto the next stage (Z-ratio -5.50) and the number of vocalisations in detour (Z-ratio -2.52) were less for older puppies.

For the Surprising Events task, latency to contact with the three items was influenced by age (latency to contact bag Z-ratio -2.16, latency to contact

umbrella Z-ratio -2.03 and latency to contact sheet Z-ratio -2.24), with younger puppies taking longer to make contact the items. Latency to approach was influenced by age for the bag (Z-ratio -2.00) and the sheet (Z-ratio -2.35), with puppies around 8 weeks of age demonstrating the quickest approach. The reapproach latency for the sheet (Z-ratio -2.52; higher younger) and the sheet approach score (Z-ratio -2.69) saw puppies of around 8 weeks approaching quickest, with older puppies contacting the sheet (Z-ratio 2.36) for the longest. The minimum distance reached for contacting the sheet (Z-ratio -2.28) was lower for older puppies, indicating closer contact with the item, and puppies around 9 weeks took the longest to reapproaching the umbrella (Z score 2.29). Vocalisations and intensity were influenced by age for the bag vocalisations (Z-ratio -2.44) and intensity (Z-ratio -2.44), for the umbrella vocalisations (Z-ratio -2.69) and intensity (Z-ratio -1.97) and for the sheet vocalisations (Z-ratio 44.72) and intensity (Z-ratio -4.70), with puppies around 8 and 10 weeks showing the lowest vocalisations. The latency to vocalisation (Z-ratio 2.46) and the vocalisation intensity (Z-ratio -2.25) for the alone time were also influenced by age, with older puppies having a lower vocalisation intensity and taking longer till first vocalising.

4.4.4 Heritability of behavioural measures

Following the removal of correlated measures (e.g., not including both total and average measures) and those with minimal variation in the data (<10% variation in data set), heritability results were ranked. A total of 79 measures remained in the data set (17 measures removed) and were ranked. See Appendix 8.10 for the measures included and removed. For all heritabilities see Table 8.25 (Appendix 8.11). The top 22 measures with heritabilities of 0.4 or above are reported below in Table 4.2. A heritability of 0.4 suggests that 40% of the variation present in the phenotype can be attributed to genetic effects.

Table 4.4 Behavioural assessment heritabilities of greater than 0.4.

Measure	Task	Heritability (h ²)	h ² S.E.
Bag minimum distance score	Surprising Events	0.83	0.14
Umbrella contact time	Surprising Events	0.68	0.13
Contact total	Novel Object	0.62	0.13
Play bow	Novel Object	0.60	0.13
Alone time – orientation to handlers	Surprising Events	0.58	0.14
Sheet response score	Surprising Events	0.58	0.14
Vocal towards door	Novel Object	0.57	0.13
Orientated towards experimenter (talking)	Human Interest	0.55	0.13
Sheet approach path	Surprising Events	0.54	0.15
Proximity to correct elbow	Odour Discrimination	0.54	0.13
Average manipulation time	Unsolvable	0.51	0.13
Average latency to solve – familiarisation	Cylinder	0.49	0.14
Alone time latency to vocalisation	Surprising Events	0.48	0.14
Bag approach path	Surprising Events	0.48	0.13
Umbrella response score	Surprising Events	0.48	0.13
Alone vocalisation	Surprising Events	0.47	0.13
Bag contact time	Surprising Events	0.47	0.13
Average interaction with Experimenter	Human Interest	0.43	0.13
Grab	Novel Object	0.41	0.12
Alone maximum vocal intensity	Surprising Events	0.41	0.13
Inhibitory score	Cylinder	0.40	0.14
Sheet contact time	Surprising Events	0.40	0.14

Using the 22 measures with heritabilities of 0.4 or above, a PCA was run to determine whether individual behaviour responses grouped together. Behaviour components were extracted for dogs with a full data set; therefore, behaviour components were available for 220 dogs. Six behaviour components were extracted (Table 4.3). Two behaviour measures (proximity to correct elbow and inhibitory score) were not present in any behaviour component.

Table 4.5 Behavioural assessment components for the top 22 heritable behavioural measures.

Behaviour measure	Task	Behaviour assessment component					
		1	2	3	4	5	6
Average manipulation time	Unsolvable						0.72
Orientated towards Experimenter (talking)	Human Interest					0.68	
Average interaction with Experimenter	Human Interest					0.58	
Proximity to correct elbow	Odour Discrimination						
Contact total	Novel Object			0.78			
Grab	Novel Object			0.74			
Play bow	Novel Object			0.62			
Vocal towards door	Novel Object			-0.51			
Bag approach path	Surprising Events		0.84				
Bag minimum distance score	Surprising Events		0.66				
Bag contact time	Surprising Events		-0.83				
Umbrella response score	Surprising Events		0.44				
Umbrella contact time	Surprising Events		-0.65				
Sheet response score	Surprising Events				0.74		
Sheet approach path	Surprising Events				0.81		
Sheet contact time	Surprising Events				-0.65		
Alone time – orientation to handlers	Surprising Events	0.73					
Alone time latency to vocalisation	Surprising Events	-0.86					
Alone vocalisation	Surprising Events	0.90					

Alone maximum vocal intensity	Surprising Events	0.91					
Inhibitory score	Cylinder						
Average latency to solve – familiarisation	Cylinder						-0.68

When reviewing the six behavioural components and the behaviours or measures they contain, the following proposed labels describe the components:

- Component 1 – Being alone post startle
- Component 2 – Bag and Umbrella startle
- Component 3 – Interaction with novel item
- Component 4 – Sheet startle
- Component 5 – Human engagement
- Component 6 – Problem solving

4.4.5 Linking 5M r-PWQ2 scores to heritable behavioural responses

Complete r-PWQ2 responses were collected for 308 of the 399 puppies (77.2%) in the sample. Puppies were either Labradors (n=121), Golden Retrievers (n=34) or crosses of these two breeds (n=153) and were from 36 sires, 69 dams and 70 litters. Mean component scores ranged from 8.38 for Anticipatory Excitement to 84.91 for Energy (Table 4.4).

Table 4.6 5M r-PWQ2 scores for puppies in behavioural assessments. Minimum, maximum, mean ($\pm$ S.E.) are shown for 10 behavioural components.

C	5M r-PWQ2 trait	Min	Max	Mean	$\pm$ S.E.
C1	Distraction and Human Focus	2.89	85.11	30.47	1.01
C2	Separation	0.00	88.80	10.71	0.97
C3	Anxiety	0.00	83.25	12.68	0.82
C4	General Excitement	0.00	99.20	29.17	1.15
C5	Attachment and Attention Seeking	0.00	82.60	29.09	0.98
C6	Impulsivity	2.60	100.00	36.72	1.22
C7	Energy	31.50	100.00	84.91	0.96

C8	Anticipatory Excitement	0.00	98.50	8.38	0.90
C9	Body Sensitivity	0.00	94.00	11.63	0.98
C10	Focus	0.00	66.00	20.59	0.83

** A lower score is more desirable and '0' indicates the dog never performs the behaviour, except for Component 7 where no score is desirable.*

Behavioural measures from the behavioural assessments were then compared to 5M r-PWQ2 scores, with C3 (Anxiety) showing the greatest number of significant correlations when compared to the other r-PWQ components (Table 4.5).

Table 4.7 5M r-PWQ2 and behavioural measures correlation for data collected during the behavioural assessments. Kendall's tau-b correlation coefficients are shown.

Behavioural measures		5M r-PWQ2 components									
Measure	Task	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Average manipulation time	Unsolvable	0.04	0.06	-0.02	0.02	0.03	0.00	0.01	0.05	0.00	0.05
Orientated towards Experimenter (talking)	Human Interest	-0.04	0.00	-0.01	0.00	-0.02	-0.04	0.05	0.01	-0.01	0.02
Average interaction with Experimenter	Human Interest	-0.08*	0.05	0.02	-0.01	0.04	-0.06	0.01	0.01	0.00	-0.02
Proximity to correct elbow	Odour Discrimination	-0.01	-0.08	-0.15**	-0.04	-0.05	0.00	0.04	0.01	0.00	0.00
Contact total	Novel Object	0.03	-0.04	-0.09*	-0.01	0.02	0.01	0.02	0.01	-0.02	-0.02
Grab	Novel Object	-0.01	-0.05	-0.13**	-0.03	0.01	-0.02	0.09	-0.04	-0.09	-0.04
Play bow	Novel Object	-0.03	-0.12*	-0.08	-0.10*	-0.10*	-0.03	0.02	0.00	-0.12*	-0.08
Vocal towards door	Novel Object	-0.03	0.02	0.01	0.06	-0.02	0.02	-0.03	-0.04	0.02	-0.01
Bag approach path	Surprising Events	-0.04	0.05	0.05	-0.02	-0.02	-0.02	-0.10*	-0.01	-0.07	0.00
Bag minimum distance score	Surprising Events	-0.02	0.02	0.07	0.03	-0.05	-0.01	-0.04	0.03	-0.04	0.02
Bag contact time	Surprising Events	0.04	-0.08*	-0.07	0.03	-0.02	0.02	0.09*	0.01	0.06	0.00
Umbrella response score	Surprising Events	-0.03	-0.02	0.10*	0.00	-0.03	0.02	-0.03	0.03	0.00	-0.05
Umbrella contact time	Surprising Events	0.02	-0.03	-0.09*	-0.02	0.01	0.02	0.09*	-0.03	0.01	-0.01
Sheet response score	Surprising Events	0.03	-0.02	0.05	0.01	0.00	0.01	-0.01	-0.02	-0.07	0.00
Sheet approach path	Surprising Events	-0.03	-0.03	0.04	-0.05	0.04	0.00	0.02	0.01	0.00	-0.06
Sheet contact time	Surprising Events	0.01	-0.03	-0.07	0.01	-0.04	-0.01	0.02	-0.01	0.04	0.01
Alone time – orientation to handlers	Surprising Events	0.00	0.05	0.08*	0.00	0.03	-0.01	0.02	0.02	-0.02	0.01

Alone time latency to vocalisation	Surprising Events	-0.01	-0.01	-0.11**	-0.01	-0.04	-0.03	-0.01	0.02	0.05	0.01
Alone vocalisation	Surprising Events	0.01	0.05	0.13**	0.02	0.05	0.02	-0.03	-0.02	-0.07	0.00
Alone maximum vocal intensity	Surprising Events	0.01	0.02	0.13**	0.03	0.03	0.03	-0.03	0.00	-0.07	0.03
Inhibitory score	Cylinder	-0.01	0.03	0.04	-0.03	-0.01	0.00	0.02	0.01	0.07	0.05
Average latency to solve – familiarisation	Cylinder	-0.04	-0.04	0.05	-0.06	-0.07	-0.04	-0.01	0.05	-0.02	-0.03

* Numbers in bold indicate a significant correlation, with one asterisk (*) indicating $P < 0.05$ and two asterisks (**) indicating $P < 0.01$.

The six behavioural components extracted from the PCA of heritable behaviours were then compared to 5M r-PWQ2 components. C3, comprising of interaction with a novel object was significantly correlated to four of the r-PWQ2 components (Separation, Anxiety, General Excitement and Body Sensitivity), with the highest correlation shown for the Anxiety component (Table 4.6). There were no other significant correlations reported.

Table 4.8 5M r-PWQ2 and behavioural assessment component correlation. Kendall's tau-b correlation coefficients are shown for 10 r-PWQ2 components and six behavioural assessment components.

5M r-PWQ2 component	Behavioural PCA component					
	1	2	3	4	5	6
	Recovery post startle	Bag and umbrella startle	Interaction with novel item	Sheet startle	Human engagement	Problem solving
C1 - Distraction and Human Focus	0.06	-0.05	-0.04	0.01	-0.07	0.05
C2 - Separation	0.05	0.06	-0.12**	-0.04	-0.05	0.01
C3 - Anxiety	0.09	0.04	-0.15**	0.04	-0.06	-0.03
C4 - General Excitement	0.08	-0.03	-0.10*	-0.03	-0.04	-0.01
C5 - Attachment and Attention Seeking	0.05	-0.04	-0.05	0.00	-0.05	0.00
C6 - Impulsivity	0.05	-0.01	-0.05	0.01	-0.06	0.00
C7 - Energy	0.02	-0.07	0.09	0.00	0.03	0.00
C8 - Anticipatory Excitement	0.01	0.04	-0.02	0.02	0.01	-0.04
C9 - Body Sensitivity	-0.06	-0.02	-0.10*	-0.01	0.03	-0.05
C10 - Focus	0.03	-0.02	-0.09	-0.05	0.01	0.01

* Numbers in bold indicate a significant correlation, with one asterisk (*) indicating $P < 0.05$ and two asterisks (**) indicating $P < 0.01$.

4.5 Discussion

The aim of this chapter was to determine whether heritable traits could be identified by evaluation of data collected during behavioural observation assessments of puppies as young as seven weeks of age, and then to compare the findings to data collected through questionnaires using the refined Puppy Walker Questionnaire version 2 (r-PWQ2; Chapter 2) at five months of age (5M). The results show that 22 of 79 behavioural measures from six tasks were shown to be highly heritable ($h^2=0.4$ or above) in the population of guide dog puppies. A PCA of the heritable behaviours extracted six components, and these were compared to the 5M r-PWQ2. The component comprising of interaction with a novel object was correlated to four of the r-PWQ2 components (Separation, Anxiety, General Excitement and Body Sensitivity). Results suggest that interaction and response to a novel item, the startle/ surprise response, problem solving, and human sociability are influenced by relationships and genetics within this population. These behaviours have the potential to impact whether a dog will be suitable as working dog (Bray *et al.*, 2021c) and therefore selecting for or against the behaviours has the potential to increase qualification rates.

This study decided to take a different approach to the analysis undertaken by (Bray *et al.*, 2020) when analysing data and heritabilities. Bray *et al.* (2020) undertook a PCA for each individual task rather than the whole assessment. For behaviour, it is common to report genetic variance and heritability estimates as groups of behaviours (often referred to as traits or components) rather than individual behaviours (Ilska *et al.*, 2017). Due to the large number of behaviour measures (96 measures in total), and a population size of 399, running a PCA was not possible for the whole data set. Bray *et al.* (2021a) undertook a PCA for individual tasks, however when looking at the behaviour measures, there were considerable similarities across tasks, so not including all behaviours in a PCA could have missed trends amongst tasks. Barnard *et al.* (2016) demonstrated issues when reviewing broad personality traits rather than individual

behaviours, suggesting that individual behaviours should be reviewed before behaviours were grouped, suggesting that the methods used in this study were appropriate for this population and sample size.

Power analysis was not undertaken in this study to determine sample size, as the project was a time bound piece of work based on GDUK resources. Post hoc analyses could have been undertaken following data collection to determine the power required to detect a significant difference. The aim was to recruit as many puppies as possible within one calendar year. The sample size was smaller than initially planned, as data was missed for puppy illness, staff holiday, logistics and weather extremes. The complexities of working with puppies were also underestimated during the initial plans. For example, the regular need to pause the assessment for toileting, or some puppies not engaging with the tasks. It was also not possible to undertake all the tasks published by Bray *et al.* (2020) due to resources and the time puppies were able to be separated from the litter as specified by GDUK. Tasks that involved no or little prior learning, and those that were relevant to GDUK were chosen as part of the assessment. The retrieval, laterality, communicate marker, arm pointing, working memory, odour control, visual discrimination and auditory discrimination tasks were excluded from the present study. There is the potential that other tasks may have shown heritabilities within this population, but it was not possible to explore all the tasks due to the above reasons. Future studies could collect data on different tasks within the cognitive battery produced by Bray *et al.* (2020).

A total of 399 puppies were used in the study, but each of the six tasks had a different number of puppies undertake or complete the task. This sample size is comparable to other studies e.g., Barnard *et al.* (2017), and greater than 168 puppies assessed by Bray *et al.* (2020) to develop the assessments. This study assessed approximately one third of the 1200 puppies bred by GDUK each year. A larger sample size or repeating elements of this assessment may have enabled a PCA to be undertaken

on all the behaviour measures. The sample size of 399 is greater than the sample size proposed by Osborne and Costello (2019), where it is suggested that for each item in the analysis, there should be 10 subjects, so greater than 60 individuals would be needed for this study if a task was considered as an item in analysis.

When analysing reactions of animals during behavioural assessments, researchers commonly use scoring systems to record the reaction of young dogs e.g., (Asher *et al.*, 2013) or quantifiable measures such as time e.g., (Bray *et al.*, 2020). The data in this study used a combination of both methods and when scoring systems were used, a clearly defined ethogram was followed to ensure consistent scoring. Scoring was undertaken mainly using video footage of the assessments, so these could be rewatched and analysed if required. It is important that researchers can agree and score consistently when reviewing behaviour (Harvey, 2021). The behavioural scoring in this study was undertaken by a small number of individuals to reduce assessor variation and the inter and intra-rater reliability is reported in the next chapter. Discussions with Dr Bray (E. Bray 2021, *personal communication*, 05 November 2021) revealed that the intra-reliability of these assessments had not previously been explored. The study could have been improved had a reliability review been undertaken in advance of the project commencing, as appropriate measures to ensure consistency could have been explored.

It was not possible to examine repeatability of behavioural responses as puppies were only at the GDNC for approximately one week. Puppies were then placed in volunteer homes around the UK, and it was not possible to follow up assessments in a standardised location. Data not presented in this thesis are being collected by GDUK to review replicability in adulthood. Approximately 40 puppies (10% of the original sample) will undertake the same assessment between 12 and 14 months of age when selected as potential breeding dogs. This is comparable to the study published by (Bray *et al.*, 2021b) where 160 assistance dogs had

assessments undertaken as puppies (between eight and 10 weeks) and then again at approximately 21 months of age. For most tasks, there was a reported increase in interaction seen with age, and individual differences were present in puppies that lasted through to adulthood. Cauchoux *et al.* (2018) also stated that cognitive performance is moderately consistent for individuals over time following the review of 44 studies on cognitive performance in 25 species. Findings from the adult assessment will review consistency in this population in the future.

The present study was based on the cognitive tasks reported by Bray *et al.* (2020) and when comparing the results from the two studies, there are many similarities. Both studies included dogs from the same breeds (Labradors, Golden Retrievers and crosses of these breeds) that were bred and raised as assistance dogs, albeit the nature of the roles was different. The sample size for the present study was larger than reported by Bray *et al.* (2020), using assistance dogs bred and trained by Canine Companions (CC), an organisation in the USA. Data for the problem-solving tasks were comparable between the populations: CC puppies on average spent 12.7 seconds manipulating the box in 'Unsolvable' and this population had an average of 11.9 seconds. Similarly, the correct first and last choice of elbows in Odour Discrimination was 54% and 73% retrospectively for CC puppies, and 55% and 79% for this population. Interestingly the eye contact was on average higher in CC puppies during Human Interest (6.3 seconds), which was nearly halved in this population at 3.3 seconds. The level of engagement post being spoken to was however higher in this population at 25.2 seconds compared to 18.5 in the CC population. The role of the working dogs and the genetics of the dogs bred for a specific purpose might explain some of these differences. The communication strategies of a CC dog would be different to guide dogs, where eye contact would not be as essential, but engagement and interaction is key as a communication strategy. Comparisons for all the tasks were not possible due to the difference in analysis. Bray *et al.* (2020)

used a PCA for Novel Object and Surprising Events, whereas this study reviewed individual behaviours.

Several studies have analysed social learning and engagement between humans and dogs (Howell and Bennett, 2011). Two of the measures related to sociability and interaction with humans had heritabilities above 0.4; engagement post talk (0.43) and orientation towards the Experimenter during talking (0.55). Bray *et al.* (2021a) also reported high heritabilities (0.43) for human interest using the same task as in the present study. Wilsson and Sundgren (1998) reported a heritability of 0.42 for social contact in eight-week-old puppies and (Van der Waaij *et al.*, 2008) reported a heritability of 0.38 for willingness to seek human contact, both studies were undertaken with German Shepherd Dog puppies. Interestingly, Van der Waaij *et al.* (2008) found no heritable aspect to seeking human contact within Labrador Retrievers (0.03 heritability) and Persson *et al.*, (2015) described heritability estimates for eye contact to be below 0.01 in a population of 498 laboratory-housed Beagles. The results may reflect the housing of these animals, and the infrequent social interaction received during their first year of life. Findings from the present study and other published work suggest that sociability in dogs may have a genetic influence in some populations, but not all.

Fear or anxiety is a negative emotional state for animals (Mendl *et al.*, 2010; Flint *et al.*, 2018b), with the behaviours reported to have high heritability estimates in dogs (Handegård *et al.*, 2023). The behaviours are also a main cause of guide dogs not qualifying (Menuge *et al.*, 2021). The present thesis has found the anxiety-based behaviours to be consistently heritable, when data was collected using owner-reported questionnaires ($h^2=0.29$ at five months of age) and in behavioural assessments. In the present study, results suggest that behaviours associated with fear were heritable (above 0.4) in puppies around seven to eight weeks of age. The Surprising Events tasks related to the startle response were shown to have the greatest heritabilities overall when compared to the other tasks,

with over half of the 22 heritabilities above 0.4 from this task. Other studies have reported fearful behaviours to be heritable e.g., Goddard and Beilharz (1983) reporting 0.46 for fearfulness in guide dogs and Ruefenacht *et al.* (2002) noting the 'reaction to gunfire' measure to have the highest heritability of all measures ($h^2=0.23$) in German Shepherd Dogs. Other factors such as adequate socialisation (Tiira and Lohi, 2015; Chaloupková *et al.*, 2018) and maternal behaviours (Bray *et al.*, 2017) have also been reported to play a key role in the development of later fear behaviours and are important to provide and monitor. However, in this population, with consistent socialisation and maternal exposure, the results suggest that selection against anxiety may be used by the breeding programme.

The flight or fight response has been discussed widely in animals. Studies of temperament or working dog suitability often include a startle related task. The present study contained two tasks that included some aspect of novelty or startle, the first being Novel Object and the second Surprising Events. Interaction with the novel object was extracted as a component within the PCA and the behaviours were shown to have high heritabilities ($h^2=0.62$ for contact with the cat and $h^2=0.60$ for presence of play bow). This component correlated to four of the r-PWQ2 components (Separation, Anxiety, General Excitement and Body Sensitivity). The component included contact time with a novel object, whether the puppy grabbed or showed a play bow and whether the puppy vocalised towards the door (i.e. where the assessors had left the room). This was similar to findings from Bray *et al.* (2020) with the first component in Novel Object referred to as 'boldness', including 'measures of approach, proximity, orientation and physical interaction with a novel object. Surprising Events had the highest the highest number of heritable items for all behavioural measures, including the contact time with the umbrella ($h^2= 0.83$) and the minimum score approached for the bag ($h^2=0.68$). While other authors have examined behaviour in puppies in response to a startle/surprise element (Alberghina *et al.*, 2020; Stolzlechner *et al.*, 2022), few report the

heritability of these responses. Findings from the present study suggest these behaviours may be influenced by breeding and thus selection against the behaviour may be possible.

Problem solving tasks are often utilised as part of cognition/behaviour studies to assess how an individual interacts with a puzzle placed in front of them. Unsolvable and Cylinder tasks involved elements of problem solving in the present study. The time spent manipulating the box in Unsolvable and latency to solve the Cylinder familiarisation stage were grouped together in a behaviour component, with heritabilities above 0.4 (0.51 for Unsolvable manipulation and 0.49 for Cylinder familiarisation latency). Interestingly, the Inhibitory control score for the Cylinder (the number of successful trials where the puppy was able to retrieve the kibble without touching the clear cylinder), was not grouped with any other behaviour in the PCA but had a heritability of 0.40. Odour Discrimination involved aspects of problem solving alongside scent detection. The time spent in proximity to the food baited elbow in Odour Discrimination was also not grouped in the PCA but had a heritability of 0.54. Studies have utilised problem solving in tasks with puppies, such as Stolzlechner *et al.* (2022) assessing 83 puppies with food under a paper cup that could be accessed if knocked over, with controlled exposure to such tasks influencing puppies ability to cope with stressors. Lazarowski *et al.* (2020) used the same cylinder task as the present study for puppies of three, six and 11-months old and found no correlations to outcome in 113 detection dogs. Findings from other studies suggest that problem solving tasks may influence behaviour but not for all working dogs. It is unknown whether these behaviours correlate to guide dog qualification in the present study.

The present study identified breed, sex and age effects on behaviours displayed by puppies, although the effects of breed and sex were minimal. Breed and sex had the greatest impact on the Surprising Events task. Age at assessment impacted all tasks, with the greatest impact on the Surprising Events tasks. Understanding the impact that age has on the

behaviours shown by puppies could aid in determining the appropriate age to undertake these types of practical assessments. Previous research (Reitz *et al.*, 2009) noted that behaviour assessments should be conducted when puppies are older than eight weeks of age due to results found in a problem-solving task. Others have suggested that the older the puppy, the better in terms of understanding their true capacities and ability to discriminate in tasks (Miletto Petrazzini *et al.*, 2020). Findings from the present study suggests that older puppies tended to perform more desirably in tasks.

In the present study it was not possible to choose breed or the age at which puppies were assessed due to this interfering with GDUK operational practices, however where possible, a day towards the end of the week stay at the GDNC was chosen so puppies were as old as possible during the assessment. Future work could enlist cohorts of different aged puppies to assess the impact of age on the early cognitive processes within puppies. The effects of maternal behaviour on puppy behaviour have been noted by researchers (Bray *et al.*, 2017) and known to impact the success of working dogs. Future studies could include maternal behaviour in analyses as a variable. Due to the reported consistency in cognitive tasks over time, the dam and sire of the puppies could have been subject to the same behavioural tasks to identify any associations between the parents and puppies. Other factors such as litter size and the impact this may have on behaviour could also be considered. In addition, the impacts of trigger stacking, when multiple stressors build on one another (Hedges, 2015), and recovery time could have been analysed during the study if behaviours from one task into the next were reviewed and the time taken between tasks recorded.

One important aspect of any assistance or working dog organisation is the outcome of the dog and whether they succeed in the working role intended. Future studies could review these behaviour measures and the link to outcome. Guide dogs tend to qualify around two years of age

following a period of approximately six months intensive training. If the heritable behaviours were also correlated with guide dog success, GDUK could consider selection and breeding towards dogs more suited for the guiding role.

4.6 Conclusion

This study has reported the heritability for a behavioural assessment in a cohort of six- to 10-week-old guide dog puppies and linked the results to owner-reported questionnaires collected at five months, answering the proposed research question. High heritabilities were identified for the fearful and anxiety-based behaviours, including startle and interaction with a novel item. Were GDUK wanting to collect this type of data in the future, undertaking a shortened version of 'Novel Object' and 'Surprising Events' task would provide valuable data and not be as intensive for data collection as all six tasks. GDUK qualification rates have fallen over the years, so should GDUK be able to utilise findings from this study and see improvements in dogs with less anxious or fearful behaviours, there is the potential to impact qualification rates in the future, thus saving the organisation resources and producing more guide dogs to support people with a vision impairment. For the wider assistance dog and pet dog population, should fearful and anxiety-based behaviours be identified using behavioural assessments or small tasks when in the nest, it may be possible to improve the presence of the behaviours in the future through selective breeding.

Chapter 5. Reliability of behavioural measures collected during a puppy assessment by human coding and machine learning methods

5.1 Introduction

Behavioural information in working dogs is commonly captured by undertaking assessments (Asher *et al.*, 2013; Harvey *et al.*, 2017; Brady *et al.*, 2018; Bray *et al.*, 2021b), or through completion of behavioural questionnaires (Serpell and Hsu, 2001; Harvey *et al.*, 2016; Bray *et al.*, 2019; Hunt *et al.*, 2020). Questionnaire-based methods enable working dog organisations to collect data quickly and at a low cost and allow for data to be gathered on behaviours observed over a period and in different situations. However, these methods can be open to levels of bias and interpretation based on assessor experience (Rayment *et al.*, 2016). Behavioural observation assessments require greater resources to undertake as they involve collecting in depth behavioural data for a series of tasks or games. Assessments have been developed to predict outcome success in working dogs (Asher *et al.*, 2013), however it is unlikely that a test can predict a dog's behaviour response in every situation (Mornement *et al.*, 2009). More recently, there has been a focus on the problem solving and cognitive abilities of working dogs (Bray *et al.*, 2020; Lazarowski *et al.*, 2020; Bray *et al.*, 2021b; Foraita *et al.*, 2021; Foraita *et al.*, 2024) and the use of physiological sensors (Marcato *et al.*, 2022) to increase the dimension of measures captured. However, any method of behavioural data collection is only useful to working dog organisations if the data collected are reliable and valid (Diederich and Giffroy, 2006; Brady *et al.*, 2018). Sinn *et al.* (2010) noted that behavioural measures should be assessed on their inter-rater reliability, test-retest reliability and predictive validity.

In comparison to questionnaires, behavioural assessments are often scored or measured by a smaller number of assessors that are

independent to the subject, providing more objective results (Tiira and Lohi, 2014). Assessments should similarly be investigated for inter and intra- observer reliability (Harvey, 2021), ensuring the data collected are consistent and valid for use in further analysis. The inter-rater reliabilities in behavioural assessments have been reported to be as high as 0.95 (Paroz *et al.*, 2008), around 0.7 (Stempiń and Strychalski, 2024) and as low as 0.44 (Alberghina *et al.*, 2020). Achieving reliability can be a significant challenge, especially in cases of testing reactions to a novel situation, as repeat testing may result in habituation (Taylor and Mills, 2006).

In a systematic review, Brady *et al.* (2018) noted that only four of the 16 papers relating to behavioural tests in working dogs had reported inter-rater reliability statistics, and concluded there was a need to improve the reporting of the reliability and validity results of published papers. During the literature review for the present thesis, few studies were identified that reported intra-rater reliability results for behavioural assessments, however a full systematic review was not undertaken. Inter and intra-rater reliability studies are often time consuming and require the results to be blinded and analysis undertaken to be undertaken by multiple assessors; potentially explaining why these are uncommon in the literature.

Guide Dogs UK (GDUK) have previously undertaken behavioural assessments over the years (see Chapter 1), and the reliability of some results have been reported. The reliability of Puppy Profiling Assessments (PPA) was not reported by Asher *et al.* (2013) or by Blythe (2020). However, the reliability of PPA was determined in a separate veterinary student project (Ralph, 2015). For the majority of tasks, inter-observer variation of 0.7 was reported and high variation in intra-observer reliability was found. Some observers agreed strongly with themselves and others showed varying levels of agreement (Ralph, 2015). Blythe (2020) reported the inter-rater reliability of CADA (Controlled Adult Dog Assessment) to be above 0.75 for all measures when two assessors that were involved in the

development of the assessment observed 10 dogs. Harvey (2014) noted the intra-rater reliability of a mobile puppy test to range from 1.00 for barking to as low as 0.31 for gaze frequency. A criticism of the cognitive assessments reported by Bray *et al.* (2020) used in Chapter 4 is the lack of reliability referred to within the paper and following discussions with the researcher, it is clear this was not undertaken.

Collecting and coding behavioural data from observational studies is often time consuming for researchers and the results are prone to subjective scoring and human error (Munch *et al.*, 2019). To improve this, there has been wide utilisation of sensors or monitors to determine objective measures such as heart rate, temperature and activity levels with an aim to understand an animal's behaviour and/ or welfare status (Kleanthous *et al.*, 2022). These devices have been shown to vary in accuracy. Kumpulainen *et al.* (2021) placed two sensors on 45 dogs, with one sensor on the collar and another on a harness, with the sensors providing accelerometer and gyroscope data. Seven activities were undertaken to assess the precision of the sensors, with activities including lying down, sitting, sniffing, standing and walking. The movement sensor on the harness returned accuracies of 91% and on the collar 75% accuracy compared to human calculations. Schork *et al.* (2023) compared results from an activity monitor to behavioural observations for 13 mixed-breed adult dogs and found significant differences between the methods, with the greatest difference of 18% for sleep at night between the collar data and human observer data. There have been differences reported between devices and variations dependent on where a device was placed (Olsen *et al.*, 2016). It is also possible that the presence of a sensor or monitor could affect the animal's behaviour, and depending on the species and the number of animals involved, the use of devices can be costly (Jukan *et al.*, 2017).

Machine learning (ML) sits within the field of artificial intelligence, and aims to automate data using computer science methods (Jordan and Mitchell,

2015). Advances and affordability of technology have resulted in wider use of ML (Congdon *et al.*, 2022), reducing the time needed for coding (Barnard *et al.*, 2016), and enabling results to be reported almost instantly. ML has been used to analyse behaviour in animals using videos, particularly animals associated with research and farming, e.g., mice (Sturman *et al.*, 2020; von Ziegler *et al.*, 2021), cows (Balasso *et al.*, 2021) and pigs (Kavlak *et al.*, 2023). Applications of ML aim to improve welfare and include predicting cattle lameness (Randall *et al.*, 2023) and the timing of calving (Borchers *et al.*, 2017), predicting body weight from feeding behaviour in pigs (He *et al.*, 2021b) and predicting the occurrence of pig tail biting events (Ollagnier *et al.*, 2023). ML has benefits as the method involves no contact/interference with animals and does not require sensors or equipment to be placed on the animal. ML can be used to monitor large groups of individuals over time and capture data with less subjective influences, for example welfare assessments (Valletta *et al.*, 2017). Using ML can also enable the extraction of many more variables when compared to human based coding methods (von Ziegler *et al.*, 2021). Developing algorithms for automatic coding using machine learning can also reduce the chance of assessor error and bias

In dogs, ML has been used alongside devices worn on dogs (Chambers *et al.*, 2021; Cleghern *et al.*, 2021; Kim and Moon, 2022), to track behaviour in behaviour veterinary clinics (Bleuer-Elsner *et al.*, 2019) and recently to track dogs' exploratory behaviour in a task (Völter *et al.*, 2023). ML has been used to predict success of working dogs in an olfactory detection programme (Eyre *et al.*, 2023) and to collate data on guide dog puppies raised by volunteers (Cleghern *et al.*, 2021). Williams *et al.* (2020) explored the use of a smart collar to gather data on guide dog puppies whilst being with puppy raisers and captured information on the incidence of barking and activity levels. The use of ML alone could result in missed or inaccurate data, for instance when an animal behaves differently to the behaviours present in the data set used to train the model. Should ML be reliable at coding behaviour in puppies, the need for human observers

would be reduced in future assessments which would improve efficiencies and standardisation of results.

This chapter features two parts: Part 1) examining the inter and intra-rater reliability of behavioural measures recorded during behavioural assessments reported in Chapter 4, and Part 2) exploring the use of ML to measure behaviour from videos of a task reported in Chapter 4 and determining correlations between ML and human-coded data.

5.2 Aims and Objectives

5.2.1 Research question

How reliable are data collected from behavioural observation assessments in puppies, and can ML be used to improve reliability?

5.2.2 Aim

To determine the inter and intra-rater reliability of a behavioural assessment for data collected by human observers and compare this to the reliability of ML.

5.2.3 Objectives

- Review the inter and intra-rater reliability of behavioural measures captured during the assessment in Chapter 4.
- Determine if ML can monitor and record puppy behaviour.
- Undertake a pilot study using ML to compare the findings of computer-generated data to data recorded by humans.

5.3 Materials and methods

5.3.1 Study design

A prospective study on a cohort of 399 GDUK puppies between the ages of six and nine weeks was undertaken in Chapter 4. Puppies were born between 08/02/2022 and 08/02/2023 and were subjected to six previously published behaviour tasks (Bray *et al.*, 2020) designed to evaluate puppy problem solving (see section 4.3.4 for behavioural task methods).

Behavioural data were collected live during assessments or coded from videos taken of the assessments using a predefined ethogram.

5.3.2 Study setting

Behavioural assessments took place at the Guide Dogs National Centre (GDNC), from April 2022 to March 2023. All assessments took place in a room with no windows on a black rubber mat with lines painted for standardisation. Two video cameras (Sony 4K FDR-AX53) were permanently housed in the assessment room, one to record a side view and the other to capture the whole testing area. See 4.3.2 for full study setting details.

5.3.3 Study animals

For the reliability study and ML pilot, puppies were born and raised by GDUK and remained at the GDNC for approximately one week where they were housed in a purpose-built puppy kennel block with littermates.

Puppies were a mixture of Labradors, Golden Retrievers, and crosses of these two breeds (Table 5.1).

5.3.3.1 Part 1- Inter and intra rater reliability.

The study population included 25 yellow puppies (8M/ 17F) from the sample of 399 puppies in Chapter 4. Puppies had completed all tasks in the assessment and were between seven and eight weeks of age.

5.3.3.2 Part 2 – Machine learning pilot

The study population included 80 puppies (45M/35F) from the sample of 399 puppies in Chapter 4 (Table 5.1).

Table 5.1 Breed and sex of puppies for reliability studies, including (a) the intra and inter-rater reliability analysis and (b) the machine learning pilot. The proportion of this subsample to the wider sample is included.

Breed proportions	Total puppies	Male puppies	Female puppies	Proportion of larger sample
(a)				
Lab	13	5	8	8.5%
Lab (0.5) x GR (0.5)	9	2	7	10.5%
Lab (0.56) x GR (0.44)	3	1	2	18.8%
(b)				
Lab	28	18	10	18.3%
GR (0.25) x Lab (0.75)	2	0	2	20.0%
GR (0.38) x Lab (0.62)	6	4	2	26.1%
GR (0.44) x Lab (0.56)	3	2	1	18.8%
GR (0.5) x Lab (0.5)	19	8	11	22.1%
GR (0.75) x Lab (0.25)	3	3	0	30.0%
GR (0.88) x Lab (0.12)	10	6	4	37.0%
GR (0.94) x Lab (0.06)	2	1	1	8.0%
GR	7	3	4	14.3%

GR=Golden Retriever, Lab=Labrador

The training set contained 30 puppies (15M/15F). Puppies were Labradors (n=13), Golden Retrievers (n=3) or crosses of these two breeds (n=14). Eight puppies were black coloured, five were golden and 17 were yellow. Puppies were from 16 litters and were an average of 6.94 weeks of age at assessment. Puppies were assessed by either Experimenter 1 (n=14) or Experimenter 2 (n=16).

The 50 puppies (30M/20F) in the testing data set were also from the same breeds (Labrador n=15, Golden Retriever n=4, Labrador and Golden Retriever crosses n=31). Puppies were from 31 litters and were either black coloured (n=13), golden (n=11) or yellow (n=24). Puppies were an average age of 7.53 weeks when assessed and were assessed by either Experimenter 1 (n=47) or Experimenter 2 (n=3).

5.3.4 Variables and data sources

5.3.4.1 *Part 1- Inter and intra rater reliability*

The 25 puppies were chosen at random by a blinded individual (a GDUK Science Team research assistant not working on the project). The individual selected seven-week-old yellow puppies that had completed all the tasks in the assessment. Using these parameters ensured that puppies looked similar in the videos and that individuals could not be easily identified. There were 46 puppies that met this criterion at the time of selection. Once the 25 puppies had been selected, video editing software (Adobe Premier Elements 2021) was used to clip the videos so no identifying information such as puppy ID and name could be seen. The research assistant allocated each puppy a number, from 1 to 25 for inter-rater reliability. Ten of the 25 individuals were then selected for inclusion in the intra-rater reliability measurement and their videos were randomly allocated a letter from A-J and a roman numeral from i to x.

Inter and intra-reliability was determined for types of behavioural measures rather than for each individual measure within each task. This was due to the time constraints on undertaking the reliability coding and the fact that many measures were similar between tasks e.g., time spent performing a behaviour, type of vocalisation etc. Videos were coded over a five-week period during August and September 2022 at set weeks apart (see Figure 5.1). All coders coded the videos in the same order and during the same weeks. Odour and Cylinder tasks were coded with no audio so

that the coders could not hear discussion from the researchers regarding correct choices.



Figure 5.1 Coding schedule for inter and intra-rater reliability video coding.

Reliability coding was undertaken by the two Experimenters in Chapter 4 (Coder 1 and Coder 2) and a lay person (Coder 3), the latter having no prior experience of dog behaviour or video coding. The same training videos were provided to the lay person as those used by the Experimenters during training. Coders were assigned a blinded number from 1-3 and after coding, results were sent to and collated by the research assistant that had selected videos. Following completion of the analysis, all data files and the dog numbers were sent to the researcher to undertake inter and inter-rater reliability analysis.

5.3.4.2 Part 2 – Machine learning pilot.

A pilot ML was explored under the support and supervision of Dr Robert Hyde (University of Nottingham). Videos from the Unsolvable task (see 4.3.4.1 *Unsolvable*) for 80 puppies were chosen and trimmed to include the first 30 seconds of the trial. Puppies were selected at random (covering a range of puppy colours and sizes). Videos were manually trimmed using Adobe Premier Elements 2021 by the same person to create a 30 second clip. The clip started when the Handler released the puppy and ended when then the Handler noted 30 seconds had lapsed and the trial stopped. The first trial was chosen for each puppy for consistency. Videos were shared with Dr Hyde on a shared drive.

5.3.5 Bias and study size

5.3.5.1 *Part 1- Inter and intra rater reliability.*

Only puppies that had completed all six tasks were included in the inter and intra- rater reliability analysis to ensure data were available to code. Puppies from the whole sample were not available as the analysis was undertaken alongside practical data collection. This was to ensure any coding errors were highlighted and picked up as soon as possible rather than once all videos had been coded. Inter and intra-rater videos were selected by an individual from the GDUK Science Team that was not involved in the coding of the videos. Researchers involved in undertaking the assessments in Chapter 4 took part in the reliability coding to ensure analysis from videos were accurate. The reliability also included an individual with no prior experience of the behavioural assessment; this individual had considerably less time to digest and process the training materials before undertaking the analysis. Therefore, there could have been a bias towards the coders involved in the original behavioural assessments and their understanding of how to code the behaviours. It was not possible to undertake analysis on all measures because some the measures, e.g. eye contact with the Experimenter is not observable from a side angle video.

5.3.5.2 *Part 2 – Machine learning pilot.*

As the puppies were recorded using two video angles (see methods in Chapter 4), sometimes the puppy was not in full view in the camera angle used. Videos were selected randomly but were checked to ensure the puppy was in view of the camera angle to ensure data were available for ML. For the ML training, the same angle was required, so some puppies were discounted. To train the model and note when the puppy was “interacting” with the box, the puppy chosen needed to demonstrate some engagement or proximity to the box. Additionally, during selection puppies that represented a range of colours, breeds and sizes were chosen to maximise variety within the ML. Subsequently there was the potential for

bias towards puppies exhibiting interactive behaviours and therefore not representing the full sample of puppies.

Due to the way the videos were recorded, every video required manual clipping using Adobe Premier Elements 2021. Study size was limited by the time available to clip videos. The sample size was initially planned as 40 videos (i.e. 10% of the sample), but this doubled during the study to enable a larger sample size to be used. It was not possible to clip data for all 399 puppies used in assessments. Sample size was also limited by the exclusion of videos clips where the angle was not optimised and therefore the puppy was not in full view during the task.

5.3.6 Quantitative variables and statistical methods

5.3.6.1 *Part 1- Inter and intra rater reliability.*

A total of 43 measures from five tasks were coded and compared during analyses (Table 5.2).

Table 5.2 Inter and intra-rater behavioural measures for data collected from five tasks during behavioural assessments.

Task	Behavioural measures included
Unsolvable	Average manipulation time, number vocalised (trials), maximum vocalisation, number on Experimenter
Odour Discrimination	First choice correct, last choice correct, sum of switches.
Novel Object	First approach, proximity total time, number of approaches, contact total, orientation total, grab, crouch, play bow, latency to vocalisation, time vocalising, vocal towards cat, vocal towards door, groan, whine, bark/yelp, howl, maximum vocal intensity, latency to contact, second approach score, second approach lure, third approach latency, third approach path.
Cylinder	Inhibitory score, inhibitory direct approach.
Surprising Events	Bag response score, bag latency to approach, bag approach score, bag approach path, bag minimum distance score, bag latency to contact, bag contact time, bag vocalisation, bag maximum vocal intensity, bag latency to reapproach, bag reapproach path, bag reapproach minimum distance score.

For inter-rater and intra-rater reliability measures, a 2-way mixed Inter Class Coefficient (ICC) for absolute agreement was used for continuous data, and Fleiss' Kappa for binary or ordinal data (see Appendix 8.12 for a breakdown of each analysis undertaken). Analyses were undertaken using R (version 4.3.1) and SPSS version 28 (SPSS Inc., Chicago, IL).

5.3.6.2 Part 2 – Machine learning pilot

SLEAP (Social LEAP Estimates Animal Poses, <https://sleap.ai/>) is an advanced labelling/ training GUI (a computer graphical user interface) installed through Python (Pereira *et al.*, 2022). SLEAP was used for extracting frames from videos and puppies using a basic skeleton were labelled. The skeleton was made up by labelling the following body parts: nose, left eye, right eye, left ear, right ear, head, neck, shoulders, front left paw, front right paw, back, back left paw, back right paw, base of tail and tip of tail (Figure 5.2).

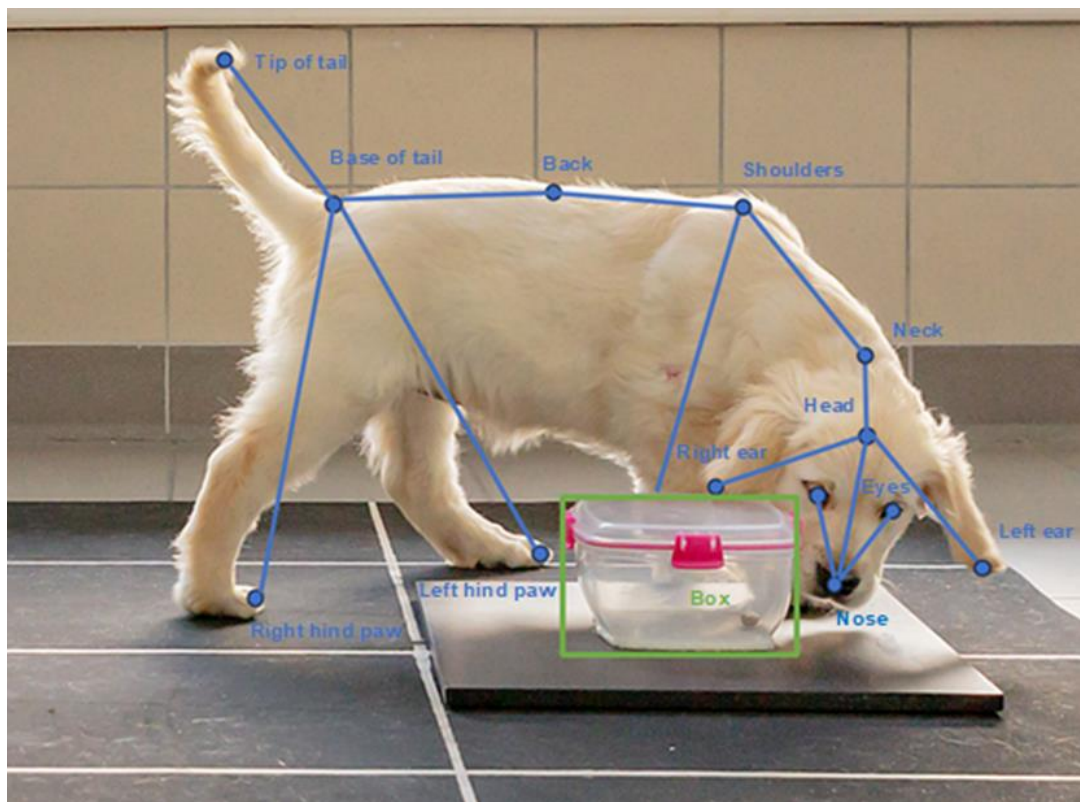


Figure 5.2 Skeleton used for machine learning. The image shows the puppy interacting with a box in the Unsolvable task and the markers placed on the puppy for machine learning.

Three scorers (BH, RH and SB) manually coded 50 videos. These were compared to the computer-generated times using Spearman's Rank correlation coefficient. Analysis was undertaken by Dr Robert Hyde. Correlations between individual scorers was undertaken by the author and SPSS used to determine the Spearman's Rank correlation coefficient.

The data were divided into training/validation (n=40) and test (n=40) datasets (Figure 5.3). Twenty frames were uniformly selected from each video, so a total of 800 frames were labelled for the 40 training/validation videos. The training/validation set were further divided into training (n=30) and validation (n=10) datasets, with a pose detection model being trained on the training set (600 labelled frames) and validated on the validation set (200 labelled frames). Once the pose estimation model had been validated, validation and test videos were watched by three human observers, and the time each animal spent interacting with the box was recorded in seconds. The pose model was then used to provide coordinates of puppies in the validation set, and was used to train a behaviour model, being tuned based on the validation set to minimise the Mean Absolute Error (MAE) in computer vs human estimation of interaction time. Once this model had been trained, it was finally tested on the external test set (n=40), which had not been utilised for training or validating either the pose, or behaviour models.

Statistical analysis was conducted using Python 3.9 (Python, 2018), with descriptive analysis being conducted using the *plotnine* package (Plotine, n.d). Mean absolute error (MAE) was calculated between computer vision predicted and mean human labelled time spent interacting with box using the *sklearn* package (Pedregosa *et al.*, 2011). Correlations between computer vision predicted and human labelled interaction time were analysed using a single raters absolute intraclass correlation coefficient within the *pingouin* package (Vallat, 2018). Interaction time was also binarized into "high" ($\geq$ median human interaction time) or "low" ($<$ median

human interaction time), and computer vision predictions of interaction time were used to predict these categories, with the accuracy of computer vision predicted interaction category being assessed using accuracy, precision and recall with the *sklearn* package.

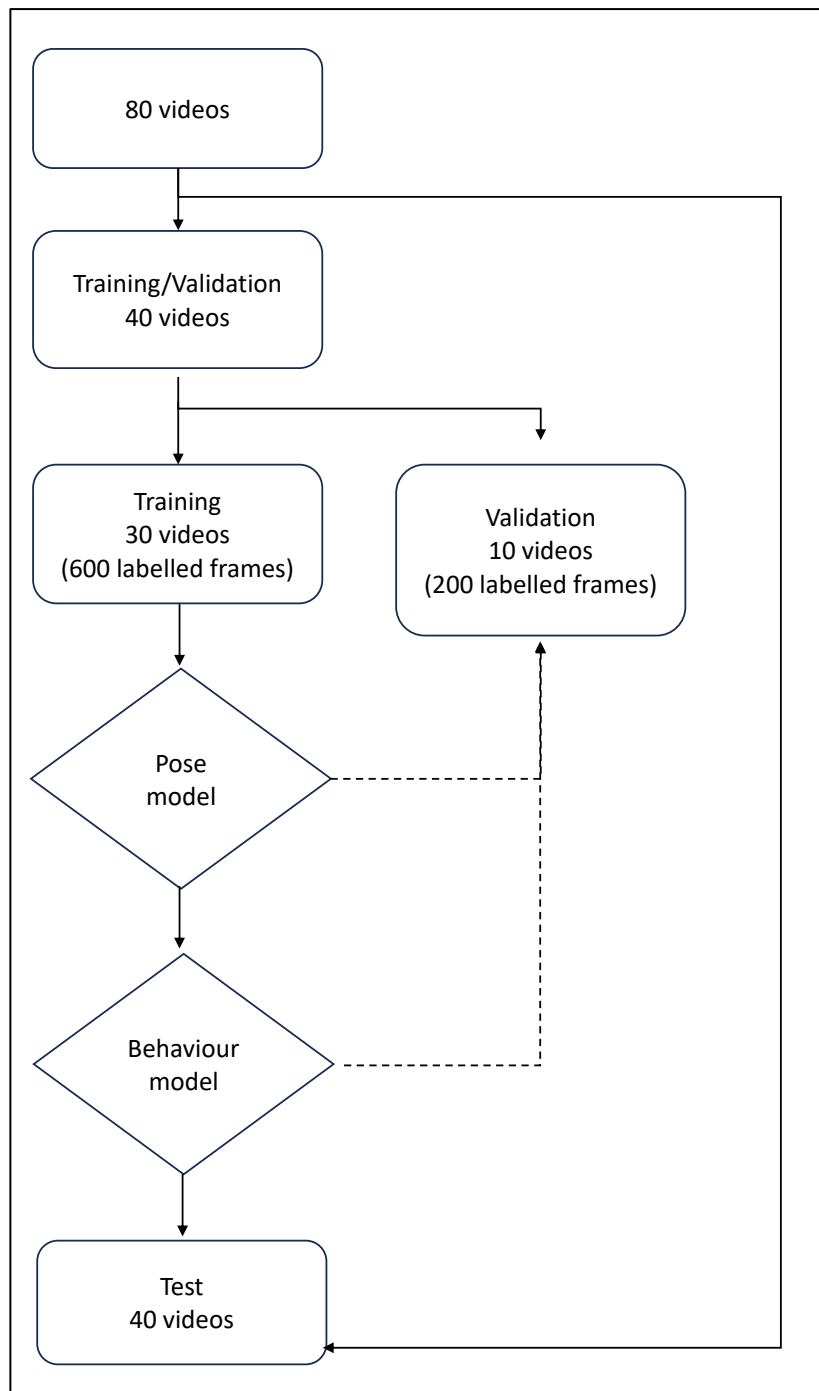


Figure 5.3 Data management pipeline for machine learning. Image produced by Dr Robert Hyde.

5.4 Results

5.4.1 Part 1 - Inter and intra-rater reliability

A total of 43 measures compared by three coders for inter and intra-rater analyses. Inter-rater reliability between assessors ranged from -0.03 to 1.00, with an average of 0.69 (Table 5.3). The intra-rater reliability for Coder 1 (C1) ranged from -0.03 to 1.00, with an average of 0.70. Coder 2 (C2) ranged from 0.55 to 1.00 with an average of 0.93. Coder 3 (C3) ranged from 0.04 to 1.00 with an average of 0.79.

Table 5.3 Inter (R1) and intra-rater (R2) reliability for behavioural measures and three coders. Scores below 0.5 reliability are highlighted and in red.

Behavioural measure	Task	R1	C1 R2	C2 R2	C3 R2
Average manipulation time	Unsolvable	0.98	1.00	0.99	1.00
Number vocalised (trials)	Unsolvable	0.93	1.00	1.00	0.84
Maximum vocalisation	Unsolvable	0.86	0.71	1.00	1.00
Number on Experimenter	Unsolvable	0.73	0.68	0.90	0.66
First choice correct	Odour	0.99	1.00	1.00	1.00
Last choice correct	Odour	0.93	0.85	1.00	1.00
Sum of switches	Odour	0.94	1.00	1.00	1.00
First approach	Novel	0.99	0.66	0.80	0.04
Proximity total time	Novel	0.97	0.98	0.99	0.99
Number of approaches	Novel	0.68	0.75	0.84	0.70
Contact total	Novel	0.98	0.98	0.99	0.98
Orientation total	Novel	0.98	0.92	0.98	0.97
Grab	Novel	0.63	0.87	0.86	0.72
Crouch	Novel	0.11	0.38	0.81	0.55
Play bow	Novel	0.69	0.72	1.00	0.71
Latency to vocalisation	Novel	0.91	1.00	1.00	0.40
Time vocalising	Novel	0.95	0.99	1.00	0.98
Vocal towards cat	Novel	0.16	0.46	1.00	0.44
Vocal towards door	Novel	0.12	0.71	0.68	1.00
Groan	Novel	-0.13	0.15	0.71	0.20
Whine	Novel	-0.03	-0.03	1.00	1.00
Bark/yelp	Novel	0.51	0.68	0.86	0.87
Howl	Novel	0.50	0.85	1.00	1.00
Maximum vocal intensity	Novel	0.54	0.89	1.00	1.00
Latency to contact	Novel	0.92	0.99	1.00	0.99
2nd approach score	Novel	0.60	0.50	0.55	0.51
2nd approach lure	Novel	1.00	1.00	1.00	1.00
3rd approach latency	Novel	0.55	0.61	0.77	0.78
3rd approach path	Novel	0.80	0.75	1.00	0.70

Inhibitory score	Cylinder	0.96	0.96	0.96	0.94
Inhibitory direct approach	Cylinder	0.39	0.44	0.63	0.60
Bag response score	Surprising	0.27	-0.07	1.00	0.46
Bag latency to approach	Surprising	0.91	0.83	1.00	0.99
Bag approach score	Surprising	0.79	0.71	1.00	0.64
Bag approach path	Surprising	0.62	0.44	1.00	1.00
Bag minimum distance score	Surprising	0.57	0.78	1.00	0.63
Bag latency to contact	Surprising	0.99	1.00	0.99	0.99
Bag contact time	Surprising	0.95	0.13	0.95	0.04
Bag vocalisation	Surprising	1.00	0.46	1.00	1.00
Bag maximum vocal intensity	Surprising	0.88	0.21	1.00	1.00
Bag latency to reapproach	Surprising	0.39	0.82	0.88	0.76
Bag reapproach path	Surprising	0.52	0.69	0.90	0.90
Bag reapproach minimum distance score	Surprising	0.72	0.60	1.00	0.90

*The Odour Discrimination task has been shortened to 'Odour', Novel Object shortened to 'Novel' and Surprising Events shortened to 'Surprising.'

For Coder 1, the maximum intra-rater reliability agreement of 1.00 (100% agreement) was present in seven of the 43 measures and an additional six measures had agreement over 90% (0.9). Therefore 13 of the 43 measures (30%) were highly correlated between the three repeated measures. Ten measures showed low agreement below 0.5.

For Coder 2, the maximum intra-rater reliability agreement of 1.00 (100% agreement) was present for 23 of the 43 measures and an additional seven measures had agreement over 90% (0.9). Therefore 30 of the 43 measures (70%) were highly correlated between the three repeated measures. No measures had an agreement below 0.5, with the minimum agreement reported as 0.55.

For Coder 3, the maximum intra-rater reliability agreement of 1.00 (100% agreement) was present for 13 of the 43 measures and an additional eight measures had agreement over 90% (0.9). Therefore 21 of the 43 measures (49%) were highly correlated between the three repeated measures. Six measures showed low agreement below 0.5.

The average intra-rater agreement was above 0.9 (90% agreement) for 17 of the 43 measures. Three measures had low agreement below 0.5 (50% agreement). The lowest three measures for intra-reliability were the bag response score and bag contact time in Surprising Events, and the presence of a groan in Novel Object. The highest average intra-rater agreement was for average manipulation time (Unsolvable), first choice and sum of switches (Odour Discrimination) and second approach lure (Novel Object). The intra-rater reliability showed 100% agreement for two measures: second approach lure (Novel Object) and bag vocalisation (Surprising Events).

5.4.2 Part 2 – Machine learning pilot

5.4.2.1 *Pose estimation*

Pose estimation models were trained for 44 epochs (number times that the learning algorithm will work through the training dataset) before early stopping, resulting in a validation loss of 0.000167 on validation frames. Behaviour model tuning resulted in the body parts: nose, front paw (left), front paw (right), left eye, right eye, left ear, right ear, head, neck and back being included, with more than one of these nodes being present within the coordinates of a box drawn. When a body part was present in the coordinates, the puppy was categorised as ‘interacting with the box’, and when none of the nodes were in the coordinates, the puppy was categorised as ‘not interacting’.

5.4.2.2 *Model validation*

The mean absolute error (MAE) over the 10 validation videos was 1.30s compared with the mean human score. The 10 validation videos were split into “high” interaction (≥ 12.0 s median human score) or “low” interaction (< 12.0 s median human score), and correctly classified nine out of 10 of these videos, resulting in an accuracy of 90%, with a precision and recall of 83% and 100% respectively.

The three scorers manually coded the interaction with the box for the 10 validation videos. The average score for Scorer 1 (S1) was 9.8 seconds for Scorer 2 (S2) 10.7 seconds and for Scorer 3 (S3) 12.6 seconds. Spearman's rank correlation coefficients were 0.85, 0.81 and 0.91 for S1, S2 and S3 respectively when compared to the model. Intraclass correlation coefficients were 0.82, 0.81 and 0.72 for S1, S2 and S3 respectively. When comparing the three scorers for the 10 validation videos, all scorers had significant agreement (Table 5.4), with S2 having the highest overall agreement between S1 (0.93) and S3 (0.95).

Table 5.4 Agreement between scorers for ML model validation for time spent interacting with the Unsolvable box during a 30second trial.

Scorer agreement	S1	S2	S3
S1		0.93**	0.87*
S2			0.95**
S3			

* $P < 0.01$ and ** $P < 0.001$

5.4.2.3 Test data set.

This model was subsequently tested on the final 40 external test videos, which were excluded from model training or validation for both pose and behaviour models. The MAE over the 40 test videos was 2.1 seconds compared to the mean human score. The 40 validation videos were split into "high" interaction (≥ 13.5 s median human score) or "low" interaction (< 13.5 s median human score), and correctly classified 38/40 of these videos, resulting in an accuracy of 95%, with a precision and recall of 91% and 100% respectively (Figure 5.4).

Spearman's rank correlation coefficients were 0.93, 0.93 and 0.89 for S1, S2 and S3 respectively when compared to the model. Intraclass correlation coefficients were 0.88, 0.90 and 0.81 for S1, S2 and S3 respectively. When comparing the three scorers between themselves for the 40 test videos, all scorers had significant agreement (Table 5.5), with all scorers agreeing > 0.97 .

Table 5.5 Agreement between scorers for ML test data set for time spent interacting with the Unsolvable box during a 30second trial.

Scorer agreement	S1	S2	S3
S1		0.97**	0.97**
S2			0.98**
S3			

** P<0.001

There were differences seen between the data reported by the model and the three scorers (Figure 5.4). The average interaction time for S1 and S2 were similar (14.3 seconds for S1 and 14.1 seconds for S2). S3 reported a higher average time spent manipulating the box (16.0 seconds). There were some videos where scorers showed high agreement e.g. 376012 Olive, where scorers were within 0.2 seconds of each other (S1 11.4 seconds, S2 11.3 seconds and S3 11.2 seconds). Other videos showed larger differences between the scorers, e.g. 375985 Garland, where S1 reported an interaction of 3.6 seconds, S2 7.2 seconds and S3 9.7 seconds, a difference of 6.2 seconds. On average there was a difference of 2.3 seconds between the maximum and minimum score for each of the 40 test videos.

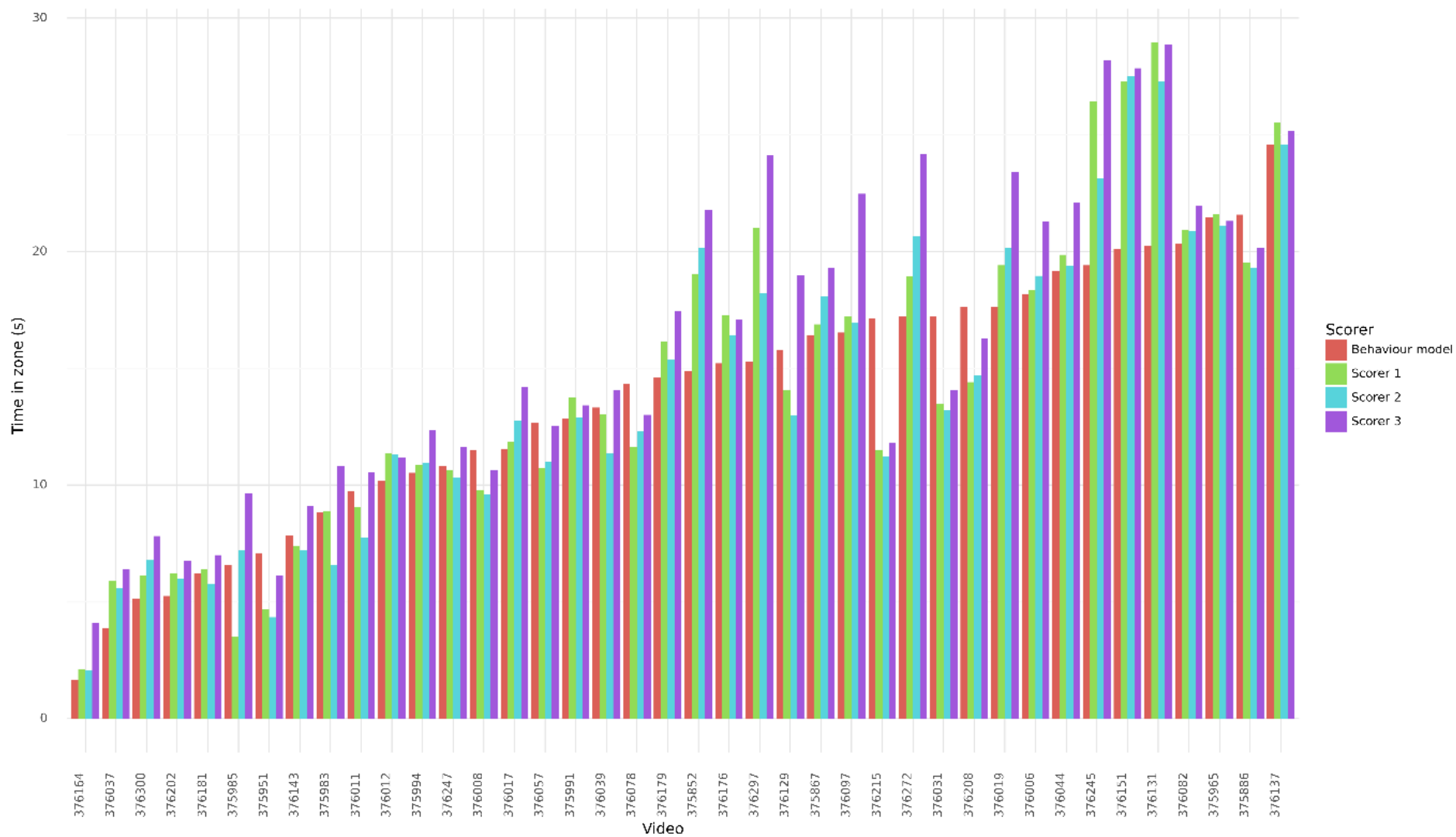


Figure 5.4 Behaviour model predictions for ML pilot for the time spent interacting with the Unsolvable box during a 30second trial. 40 videos were compared to human scorers.

5.5 Discussion

This study has reported variation in reliability (both inter and intra rater reliability) of behavioural measures and demonstrated that computer vision can be used to monitor behavioural tasks in puppies. As noted by Sinn *et al.* (2010), behavioural assessments should be judged on inter-rater reliability, test-retest reliability and predictive validity. It is therefore important that the reliability of measures is reviewed during any assessment. Bray *et al.* (2020) reported the cognitive abilities of assistance dog puppies but failed to report the reliability of the measures. The present study reported an average inter-rater reliability of 0.69 for the 43 measures analysed, indicating moderate agreement (stated as 0.60 to 0.79 (McHugh, 2012). Intra-rater reliability showed differences between the three coders, with one coder having an average agreement of 0.93 (Experimenter 1), one of 0.79 (lay person with no prior knowledge for assessment) and another of 0.70 (Experimenter 2), highlighting variances between individuals. Computer vision was able to correctly classify puppies interacting with the Unsolvable box with an accuracy of 95% in an unseen test set of 40 videos compared with human classification. This algorithm could be utilised to objectively and continuously monitor animal behaviour based on time spent in zones, circumventing the time consuming and costly current requirement for human monitoring.

Both Part 1 and Part 2 of this study included a mixture of researchers involved in the behavioural assessments and researchers with no prior knowledge of the assessments. One coder was the same between Part 1 and Part 2, so for consistency it would have been interesting to compare the same three coders between the different methods, however this was not possible with resources and availability at the time of the study. Part 2 resulted in the findings from Part 1, so the two studies were not planned in parallel. Both Parts 1 and 2 could have also included a greater number of the behavioural measures, more videos and scores from a larger number of assessors for comparison. Due to time constraints and the time it took

to code videos for one puppy, the behaviours or timings examined were selected to cover each type of measure, but not every instance of the measure.

Following the outcomes of Part 1, many videos had to be re-coded for use in Chapter 4. Initially both assessors that were involved in handling and assessing the puppies in Chapter 4 were coding videos. Following discussions with Dr Bray during study development, Dr Bray suggested that the tasks be divided between individuals, so the same person was coding the same task for consistency. As Coder 1 was shown to be less reliable for some tasks, it was decided that all the videos should be re-coded by the most reliable coder to ensure data were as accurate as possible for heritability analysis. This caused a significant delay in when data were available and increased the workload of Coder 2. This could have been avoided had the intra-reliability analysis been undertaken in advance of the videos being coded and the provision of more training and monitoring given to Coder 1.

Part 2 only focussed on one task and one specific measure (Unsolvable - time spent interacting with the box). The study was a small pilot investigation due to limited time available and reliance on the expertise of another researcher at the University of Nottingham. The study could have been improved if more tasks or behaviours were included to explore ML on a variety of different measures. The results from Part 1 showed the average manipulation time interacting with the box to have one of the highest reliability scores of all measures, (1.00-0.99 inter-rater reliability and 0.98 intra-rater reliability). As such this measure was chosen to use for the ML pilot, alongside the fact that this was one of the simpler measures to train for ML. The ML model generally performed well, although there were several outliers where predicted interaction time was significantly different to human scorers. For example, in one video (puppy ID 376215; Figure 5.4), computer vision was significantly higher than all three observers, due to the puppy placing the front paws just behind the

box, meaning that whilst human observers could interpret that the animal was “not interacting” with the box, the computer vision model could not account for the perspective and intention of the animal and classified these frames as “interacting”. There is a limitation of ML being able to distinguish and recognise an animal’s intent, which is worth considering for analysis of behavioural assessments. Standardising and choosing an appropriate camera angle (e.g. top-down camera placement) may avoid some of these errors. The camera angle for data captured during the assessments was not considered with ML in mind and could be improved in future studies. Yet when comparing machine learning to human assessors, there will also be limitations, such as the human eye missing details and not capturing the same frame rate capabilities as ML.

The inter and intra-rater reliability results reported in the present study are comparable to other studies. The reliability of a behavioural test for 105 seven-week-old puppies assessed by three assessors for five measures showed inter-rater reliability to range from 0.44 to 0.70, with an average of 0.59 (Alberghina *et al.*, 2020). Whilst that study had a larger sample than the present study, there were considerably less measures, and a 3-5 point scale was used to measure behaviour, so there was less variation in the data variables compared to the present study. Harvey *et al.* (2016) reported the intra-rater reliability of a puppy test to vary from 0.35 for licking behaviour to 1.0 for shaking behaviour for 40 guide dog puppies assessed at five and eight months of age. The intra-rater reliability of lip licks was low, suggesting that more subtle behavioural displays or behaviours open to interpretation, such as ‘groan’ average intra-reliability of 0.35 in the present study, are less reliable. The average inter-rater reliability for a military working dogs assessment was 0.77 for 15 items (Sinn *et al.*, 2010). The present study included a greater number of measures (n=43), and therefore coders had to focus on more behaviours which could have affected the reliability due to assessor fatigue. Fatigue has been shown to increase the chance of errors (Dawson *et al.*, 2012) and tiredness has been linked to a reduction in cognitive function (Lim and

Dinges, 2010). It is therefore important for researchers to have breaks when coding videos to reduce unintentional errors.

Assessing the behaviour of dogs or puppies requires a level of skill and training (Mornement *et al.*, 2009). This study reported some differences between coders, some having more experience in assessing dog behaviour. The same coder was shown to be the most reliable in both Parts 1 and 2. The coder was the author of this thesis and was trained first hand by the developer of the assessments (Dr Emily Bray). It could be argued that the coder was more reliable due to having spent the longest time involved in the assessment when compared to the other coder and scorers. Yet having experience does not mean that individuals with the same level of experience will agree with one another or be able to recognise the same behaviours displayed. The Dogs Trust reviewed the reliability of 17 assessors with formal training and more than eight years' experience working with dogs, and the agreement varied depending on the behaviour measures, with agreements as low as 0.03 reported (Diesel *et al.*, 2008), highlighting the considerable variation that can be present when assessing dog behaviour.

Another factor to consider when measuring behaviour is the reaction times of an individual. A study reviewing a change of direction in 62 basketball players found that three scorers reported significant differences between each other when using stopwatches, with one scorer being significantly faster than the other two scorers (Chen *et al.*, 2021b). Part 2 of the present study found one scorer to have consistently higher timing scores for interaction with the box than the other two scorers, highlighting how individual variation will be present when using stopwatches. For the 40 videos in the ML test data set, the mean time difference between the model and the scorers was 2.1 seconds, which for a 30 second trial equates to around 6% of the trial. However, as the trial was short, there is a need to account for the reaction time of the assessors to click start and stop on the stopwatch, and this could account for some of the differences reported. ML can precisely time interaction from the immediate moment

the puppy engages with the box, but there will be a lag time to react for scorers to start a stopwatch. When measuring the start and stop time of reading 12 sentences, the reaction times of individuals with stopwatches were shorter than an automated computer programme timer for approximately 67% of measures (Radner *et al.*, 2017), and individuals were only required to press the stop watch once. The Unsolvable task in the present study required coders to start and stop the stopwatch multiple times during the 30 second trial depending on how many times the puppy interacted with the box. The error between the computer-generated time and the assessor recorded time is likely to be higher. It is also important that coders use the same equipment due to potential differences between stopwatches (and whether these were on a mobile device or not). The scorers in Part 2 did not all use a stopwatch which may have also inflated differences between coders. The coders in Part 1 were all given the same make and model of stopwatch, but the devices used were different as the coding was taking place at the same time.

The results in the present study demonstrate the importance of reliability for behavioural measures and the need to undertake inter- and intra-rater reliability studies. Unreliable data will affect data interpretation and could influence any further findings, such as reviewing the heritability of behaviours. There is also the need to check assessors, even when training has been undertaken, as individual bias and errors need to be identified early to reduce the need to reanalyse data. Repeated training or refreshing of information enhances retention (Wiklund-Hörnqvist *et al.*, 2014). There is the potential for a drift in coding over time, so it is important that refresher training is undertaken.

It is unknown how the findings from Part 2 of this study would generalise to other animals, other breeds of dog or data collected in different environments. Future studies could explore the use of ML for more of the tasks in Chapter 4 or the use of more readily available video footage, such as CCTV footage. GDUK has CCTV cameras in all kennel blocks at the

GDNC, so there is the possibility of using video footage of puppies or adult dogs to monitor behaviour using ML in the future. Further studies may be able to utilise genotype data collected from Born to Guide (a study undertaken by GDUK to collect DNA samples from around 3000 dogs) and relate findings to behavioural measurements using ML. In addition, when the outcome (whether the puppy qualified as a guide dog or was withdrawn) of the individuals analysed in Part 2 is known, ML could be used to compare data for qualified vs withdrawn dogs and determine whether the computer can identify differences between individuals that human coders may not necessarily pick up or notice when reviewing videos. It may therefore be possible for computer vision to predict the ultimate success or failure of assistance dogs using videos of behaviour.

5.6 Conclusion

Observational behavioural studies require a great deal of time and resources which is particularly important to assistance dog organisations that rely on charitable funding as the primary source of income. Therefore, it is important to ensure that such projects are value for money by providing insightful and reliable results to inform future decision making. The present study answered the research question and identified variation between coders that further supports the importance of reviewing the reliability of behavioural measures. Despite the coders all receiving the same training videos and the same time allocation to undertake analysis, the variation between individuals was clear. An emphasis on regular checking and reviewing of data is important to ensure the success of any behavioural study. Findings from the present study suggest that ML could be a useful tool for researchers to consider improving the reliability of coding for behavioural assessments. A multidisciplinary approach to behavioural analysis would enable computer scientists to work with animal behaviourists, reducing the time required for analysis and improving data consistency and standardisation. There is the potential to further explore the use of this technology to enhance behavioural data collection in

different situations, including assistance dog organisations, veterinary practices, and rehoming shelters.

Chapter 6. An exploratory genome-wide analysis of r-PWQ2 components at eight and 12 months

6.1 Introduction

The field of genetics aims to understand the processes by which characteristics are inherited. Genetic studies utilise DNA information in the genome (Dodgson *et al.*, 1997; Vignal *et al.*, 2002), utilising methods such as whole genome sequencing to explore the genetic code of an individual (Ng and Kirkness, 2010). Single nucleotide polymorphisms (SNPs) are single base pair changes in the DNA sequence (Marth *et al.*, 1999), which are used to identify genome regions associated with complex traits or disease (Shasstry, 2007) and in humans, SNPs occur approximately once in every 300 base pairs of a sequence (Reich *et al.*, 2003). Variation is referred to as polymorphism, which occurs due to mutations in DNA, and is thought to occur in around 0.1% of loci between individuals (Shasstry, 2007). There are thought to be around 19,000 dog genes (fewer than in humans), with approximately 94% of the dog genome showing synteny (genome similarities) relative to the human genome (Ostrander and Wayne, 2005). Due to advances in genetic technologies, scientists have a greater understanding of the genes involved in some complex and common health conditions and behavioural characteristics.

In 2004, the first dog draft sequence genome (a Boxer dog called Tasha) was made publicly available (Lindblad-Toh *et al.*, 2005), commonly referred to as the 'reference sequence'. This information has enabled researchers over the past two decades to undertake studies of the dog genome by determining variations from the reference sequence. Since the sequencing of Tasha, advancements in the field of dog genetics include improvements in sampling methods and the number of SNPs present on genotyping arrays (Van Steenbeek *et al.*, 2016). The dog genome has been studied and updated by scientists, with the reference Boxer Tasha now sequenced to over 2.3 million units in length (Jagannathan *et al.*,

2021), and other whole genome sequences available for different dog breeds such as Labrador Retrievers (NCBI, 2020).

SNPs are commonly found within a gene or in a regulatory region near a gene, affecting gene function, which can lead to disease (Kaur *et al.*, 2019). SNPs associated with specific and rare phenotypes may be located in uncommonly sequenced areas of the genome, so understanding the genes involved becomes challenging (Nuttall, 2013). It has been suggested that future advancements from SNP and polymorphism analyses could determine the susceptibility of an individual to many common diseases and individualise medications provision (Shastry, 2007). Studies utilising SNP arrays are referred to as genome-wide association studies (GWAS), with the aim to review genetic variations present in the genome to identify genotype-phenotype associations (Ball, 2013; Dreger *et al.*, 2016; Tam *et al.*, 2019). However, a GWAS will not identify specific causal mutations on a trait or disease unless the loci at which those mutations occur have been included in the SNP array. With ever increasing advancements and the ability to sequence more SNPs (as demonstrated with the increased genome coverage for the Boxer Tasha), in the future there may be the ability to study more complex or rare disorders that occur in uncommon gene mutations.

A high quality reference genome (Wang *et al.*, 2021), affordable SNP genotyping and ease of collecting DNA data (e.g. non-invasive saliva swabs) have contributed to the dog being used as model species for complex traits and diseases in humans (Ke *et al.*, 2011; Van Steenbeek *et al.*, 2016). Mellersh (2008) observes the dog to be an excellent model for human disease due to; similarities in the pathology of diseases and species physiology (more so than other laboratory animals); the wealth of data available on canine diseases and the close shared environment of dogs/humans following domestication. It has been suggested that of the 858 genetic traits and disorders in the dog, 420 are potential models of human disease (Sydney, 2023). Within behaviour, correlations have also

been found between human and dog traits or characteristics. Persson *et al.* (2016) found that SNPs in the dog associated with “human contact seeking” had previously been associated with attention deficit hyperactivity disorder (ADHD) in humans. A study looking at obsessive compulsive disorder (OCD)-like tendencies in dogs using continuous circling as a representative phenotype for OCD in Belgian Malinois and Kunming Dogs found differences in 11 candidate genes that were also associated with OCD in humans (Cao *et al.*, 2021). Another benefit of using the dog genome rather than the human genome, is that far fewer SNPs are required for a dog analysis (approximately 10,000 SNPs) than in humans (approximately 500,000 SNPs), due to the extent of linkage disequilibrium in dog breeds resulting from differences in population structure. Extended regions of linkage disequilibrium mean associations can be detected over greater genomic distances.

The heritability estimate is the proportion of phenotypic variance explained by the variance of additive genetic effects (Zhu and Zhou, 2020), and the heritability can be either pedigree or genomic-based. Genomic data is reported to deliver more accurate estimates of genetic influences than pedigree estimates alone (Speed *et al.*, 2012; Keijser *et al.*, 2018a), due to more precision in relationships between individuals and locational precision in chromosome regions (Gibson, 2018). Pedigree-based heritability estimates evaluate heritability without requiring information on the inheritance of individual loci as correlations are inferred from family pedigrees (Xia *et al.*, 2016). Pedigree-based heritabilities rely on knowing the pedigree history of individuals and having phenotypes available for related individuals. As such, studies commonly utilise Kennel Club registered dogs where the pedigree is known. Genomic-based heritabilities enable comparisons between unrelated individuals and can be undertaken without the constraints of a pedigree because associations with genetic markers can be identified using DNA (Speed *et al.*, 2012). The heritability estimates found by pedigree and genomic-based methods can therefore differ. Although examining species of trees, an interesting

finding by Calleja-Rodriguez *et al.* (2020) found that prediction accuracies in genomic models were between 50 and 143% higher than pedigree-based selection. Similarly, Veerkamp *et al.* (2011) found that heritability and genetic correlations were more accurate with genomic relationships compared with using pedigree relationships in a population of Holstein-Friesian heifers.

For genetic studies to be useful and reliable, the phenotypic information needs to be accurate and abundant (Liinamo *et al.*, 2007; Nuttall, 2013; Ilska *et al.*, 2017; Friedrich *et al.*, 2019). To undertake a reliable quantitative genetic analysis study, it is implied that at least 500-1000 dogs with known phenotypes be included, and that all descendants of the sire and dam be studied where possible (Liinamo *et al.*, 2007). It is also essential that parentage is recorded accurately in the pedigree data to infer correct genetic relationships (Lewis, 2019). However, due to sequencing costs it is not always possible to have large samples for genomic studies. Nevertheless, studies have reported GWAS findings from as few as 104 dogs (Bellamy *et al.*, 2018).

GWAS in dogs have detected associations with morphological (Plassais *et al.*, 2019; Sheet *et al.*, 2021), reproductive (Smith *et al.*, 2019), health (Forman *et al.*, 2012) and behaviour traits (Sarviaho *et al.*, 2019; Fallahi *et al.*, 2024). For behaviour, Sarviaho *et al.* (2019) discovered significant loci on Chromosomes 20 and 7 associated with canine anxiety and fear in a cohort of 330 German Shepherd Dogs. Fallahi *et al.* (2024) found two regulatory SNPs in the TAC1 gene to be associated with excitability and trainability in 123 male Belgian Malinois dogs. Both these studies have focussed on specific populations and behaviours, rather than exploring more breeds and a wider repertoire of behaviour. Therefore, the replicability and applicability to the wider dog population is limited. Breeding for behaviour can often be more challenging than breeding for physical traits (Plomin, 1990; Evans *et al.*, 2015) due to data collection methods and subjectivity of measures. Breeding for behaviour can

improve animal welfare by selecting animals with more desirable and less abnormal behavioural displays. As previously discussed in Chapter 1, specifically for working dogs, understanding the genetic basis for behavioural displays could be of significant value in producing dogs with more suitable working behaviours. Following analysis of the pedigree-based heritability estimates using the r-PWQ2 (Chapter 3), results suggest a genetic influence for some behavioural components. Behaviours associated with excitement, anxiety, separation, energy and body sensitivity were suggested to have a heritable component. Genomic-based heritabilities would provide refined estimates and determine the viability of genomic breeding values (GBVs) and GWAS could potentially reveal specific gene effects on behaviour and so contribute to a deeper biological understanding of their aetiology.

6.2 Aims and Objectives

6.2.1 Research question

Are genomic heritability estimates calculated from genotyping animals comparable to the pedigree heritability estimates calculated in Chapter 3?

6.2.2 Aim

To undertake genetic analyses on a sample of 96 genotyped dogs with r-PWQ2 data collected at eight and 12 months of age.

6.2.3 Objectives

- To collect saliva samples from 150 dogs, representing a range of sires and dams within the GDUK breeding programme.
- To select 96 samples for 230K SNP genotyping.
- To undertake genomic-based REML (restricted maximum likelihood) and GWAS studies on the genotype data, taking into consideration

other confounding and environmental variables such as puppy raiser home and experiences.

- To compare genomic-based heritability estimates to pedigree-based estimates from Chapter 3.
- To identify genome regions and putative candidate genes associated with significant genomic-based heritability estimates and explore links to published behavioural phenotypes.

6.3 Materials and methods

6.3.1 Study design

This study involved collection of saliva swabs, and a mixture of both retrospective and prospective methods to collect behaviour questionnaires for GDUK dogs born over four years (2016-2019; see full methodology for questionnaire collection in Chapter 2). Saliva swabs were received from 120 dogs and 96 dogs from the sample were selected for genotyping with a high-density SNP genotyping array. Data collected from the r-PWQ2 at eight (8M) and 12 months (12M) were analysed. Ethical approval was granted by the University of Nottingham Veterinary School's Committee for Animal Research and Ethics (reference 3248 200108).

6.3.2 Study setting

All dogs were bred by GDUK between 2016 and 2019 and placed into puppy raising (see Chapter 2; 2.3.2 Study setting).

6.3.3 Study animals

The dogs in this data set were a subset of subjects from heritability analyses in Chapter 3. The larger data set was filtered for the three most represented breeds (Golden Retriever x Labrador (n=428), Labrador

(n=301) and Labrador x Golden Retriever (n=224)) with completed 12M r-PWQ2 data (n=953). Where possible, two offspring per dam were randomly selected from the data to create a genotype sample set (selection performed by Dr S. Blott). Three hundred and forty-three dogs were randomly selected from Golden Retriever x Labrador (n=428), Labrador (n=301) and Labrador x Golden Retriever (n=224) dogs with completed 12M data (n=953), representing 61 sires and 181 dams, which represented 70% of GDUK breeding stock (n=345) at the time of sampling. Two offspring from each dam were selected where possible, but for 19 dams only one offspring was selected.

Owner consent to be contacted for research purposes was determined, and dogs of those that had opted out were removed from the data. Following the contact checks, further filtering was undertaken, ensuring offspring with unique dam IDs were chosen where possible and that all sires (n=154) were represented. Where two dogs from a sire and dam (full siblings) were available, one dog was chosen at random. The distribution of breed, sex and behaviour data were examined to ensure dogs selected were representative of the sample. Approval for contact was sought from GDUK staff members responsible for the dogs (see Figure 8.5 in Appendix 8.13), resulting in two further dogs being withdrawn from the data set.

Following checks for consent, contact and owner acceptance to take part in the study, 120 swabs were returned. Ninety-six dogs were selected for genotype analysis based on sire and dam representation and 12M r-PWQ2 scores to represent a spread of behavioural phenotype whilst ensuring the maximum number of dams and sires were represented. Dogs genotyped were Golden Retriever x Labrador (n=47), Labrador (n=34) and Labrador x Golden Retriever (n=15) and 51 were female and 45 males. Dogs were born between 2016 and 2019, with two dogs being born in 2016, 44 born in 2017, 43 born in 2018 and seven born in 2019. The dogs represented a range of stages within GDUK at the time of sampling, including breeding stock (n=8), buddy dog in training/qualified (n=3),

puppy raising (n=2), on-class (n=1), training (n=25), qualified guide dog (n=27) and withdrawn (n=30).

The 8M REML heritability analyses were undertaken on 81 dogs. The 81 dogs (38 male, 43 female) represented 47 sires, 81 dams and were from one of three breeds Golden Retriever x Labrador (n=39), Labrador (n=28) and Labrador x Golden Retriever (n=14). Dogs were born between 2016 and 2019, with two dogs being born in 2016, 38 born in 2017, 35 born in 2018 and six born in 2019. Due to the small sample size at both 8M and 12M, basic heritability models were run including only sex and breed as fixed effects.

6.3.4 Variables and data sources

6.3.4.1 *Phenotype and individual data*

Phenotype data were collected using the r-PWQ2 behaviour questionnaire collected at 8M and 12M as previously described in Chapter 3. Information including sex, breed, GDUK staff contact, the owner responsible for the dog, sire ID and dam ID and r-PWQ2 data was also collated. The r-PWQ2 data consisted of 39 unique questions which produced 10 behavioural component scores in addition to information about the puppy raisers home environment. Component numbers were abbreviated so Component 1 is referred to as C1, Component 2 as C2 etc.

6.3.4.2 *Genotype data*

DNA was collected using DNA Genotek Performagene saliva swabs (<https://www.dnagenotek.com/us/products/collection-animals/performagene/PG-100.html>).

The owners of dogs (whether puppy raisers, boarders, breeding dog holders, clients or rehoming) were contacted directly by their preferred method and asked to participate in the study. If owners agreed, a letter

explaining the study and sampling process (see Figure 8.6 in Appendix 8.13 for a copy of the letter), sample instructions (provided by Performagene; see Figure 8.7 in Appendix 8.13 for a copy of the instructions), a sealed saliva swab and a prepaid addressed envelope to return the sample to GDUK was sent. Returned samples were stored at room temperature as per manufactures instructions.

Genotype data were obtained using the Illumina Canine High Density Beadchip, containing 230,000 SNP markers derived from the CanFam3.0 assembly with DNA extracted from saliva swabs. Genotyping was undertaken by NEOGEN, Ayr, Scotland, UK.

6.3.5 Bias and study size

Study size was limited by the funding available for genotyping and the dogs available with r-PWQ2 data at 12M. Sample collection was limited to owners that consented to be contacted for research through GDUK's databases and for owners that replied to the research request. For qualified guide dogs, owners with sighted support were preferentially chosen due to the complexity of sample collection and accessibility of instructions and labelling samples, as such this removed the ability to sample some dogs. It was not possible to match the number of pure and cross breeds in the sample for genotyping when also accounting for the issues with sampling and the need to represent different sires and dams from the population. Therefore, the number of purebreds and crossbreeds was not equal in the sample, and it was not possible to review differences between these due to the small sample of purebred Labradors. Other GDUK breeds/breed mixes e.g. German Shepherd Dogs, were not used in the genotyping study, however analyses of closer related breeds improves the chances of more meaningful results. Similarly due to the small sample size, it was only possible to run simple GWAS models and all factors that had been shown to impact r-PWQ2 scores from Chapters 2 and 3 could not be included in the final GWAS model.

It was not possible to use 5M r-PWQ2 data as the funding for sampling and had to be spent before a full 5M data set was available. Due to the Covid-19 pandemic, data collection for questionnaires ceased due to the impact of the pandemic on human attitudes and dog behaviour. There could also be variation in the phenotypes used for the GWAS due to the type of data being collected and owner biases in the reporting of dog behaviour. Puppy raisers may not want to report negative or undesirable behaviours seen as they may feel it would reflect badly on them as a guide dog volunteer.

6.3.6 Quantitative variables and statistical methods

6.3.6.1 *Genotyping quality control procedures*

The data were subjected to a number of quality control (QC) procedures. These were performed using PLINK v1.07 (Purcell *et al.*, 2007). QC procedures included sex verification, SNP missingness, extreme heterozygosity identification, and identity-by-state (IBS) or relatedness evaluation. Details of the commands used for each QC procedure are shown below, with a description of the command. The command “--file” is used to open the PED/MAP files, containing the genotyping and pedigree information for each individual and “--out”, is used to specify the name to be given to the output file.

For Step 1, a sex check was undertaken, using the command “*plink --noweb --dog --file PBQgenotype --impute-sex --out PBQgenotype.sexcheck*”. The sexes assigned to the animals were verified using calculated X chromosome inbreeding coefficient (F), where $F > 0.75$ were classified as male and $F < 0.55$ female. The PLINK default F thresholds are $F < 0.2$ for females and $F > 0.8$ for males but were reviewed for this population and altered accordingly ($F > 0.7$ males and $F < 0.6$ females).

Step 2 calculated missing SNP data rates for individuals using the command “*plink --noweb --dog --file PBQgenotype --missing --out PBQgenotype*”. SNP missingness returned the location (locus), number and proportion of missing SNPs for an individual (i.e. the SNPs call rate).

Step 3 determined the observed rate of heterozygosity per individual using the command “*plink --noweb --dog --file PBQgenotype --het --out PBQgenotype*”. The individual observed heterozygosity was calculated using the equation:

$$\text{Observed Heterozygosity} = \frac{N(NM) - O(HOM)}{N(NM)}$$

Where $N(NM)$ = number of non-missing genotypes per individual, and $O(HOM)$ = observed number of homozygous genotypes, as given by PLINK output files. From this, mean observed heterozygosity and standard deviations were calculated. Individuals with high heterozygosity were reviewed, since this can indicate poor specificity of genotyping (the DNA not fully hybridizing to chip). No individuals had an observed heterozygosity below or above three standard deviations from the mean and therefore no individuals were excluded. Results were reported as mean $\pm$ standard deviation.

Step 4 identified any duplicated or related individuals. SNPs were pruned so that for correlated SNPs (those in high linkage disequilibrium (LD) with each other) only one was retained in the analysis set. The command “*plink --noweb --dog --file PBQ genotype --indep-pairwise 50 5 0.8*” was used. The command *--indep-pairwise* allows LD pruning based on pairwise r^2 values. In this command the 50 indicates a window size of 50 SNPs, 5 indicates that there are five SNP intervals, and the r^2 threshold is set as 0.8. LD-based SNP pruning created the files “*plink.prune.in*” (SNPs to include) and “*plink.prune.out*” (SNPs to exclude).

Step 5 identified identity-by-state (IBS) distances between individuals which were indicative of relatedness using the command “*plink --noweb --dog --file PBQgenotype --extract plink.prune.in --genome --out PBQgenotype*”. Individuals were then plotted on a scatterplot to visualise any population stratification. All individual pairs had an IBS<0.5 (where IBS=0.5 is indicative of a parent/offspring or sibling relationship) and therefore no individuals were excluded. This was expected for individuals as dogs were selected to represent as many sires and dams as possible in the sample. Results are reported as mean $\pm$ standard deviation.

The last step (Step 6) of the QC procedures was to identify SNPs with an excessive missing data rate across all individuals using the output file “*PBQgenotype.lmiss*” which was generated in Step 2. The results enabled filter thresholds for missing data to be determined. The missing rate per SNP command was run as “*plink --file PBQgenotype --geno 0.1*” indicating SNPs with a genotyping call rate of less than 90% would be excluded. The missing rate per individual was run as “*plink --file PBQgenotype --mind 0.05*” indicating that any individual with less than 95% genotypes present would be excluded. The minor allele frequency command was run as “*plink --file PBQgenotype --maf 0.01*”, indicating that SNPs with a minor allele frequency of less than 0.01 (1%) would be excluded from the analysis. Lastly, the Hardy-Weinberg Equilibrium (HWE) test was run as “*plink --file PBQgenotype --hwe 0.001*”. This ensured that SNPs with genotype frequencies/proportions that did not fulfil HWE were excluded. HWE can be used in GWAS to estimate the number of homozygous and heterozygous variants allele frequencies in future generations (Duhovnikoff and Leventhal, 2016; Abramovs *et al.*, 2020), indicating the presence of genetic drift in populations. The QC steps resulted in no individuals being excluded from the data set for genome analyses.

6.3.6.2 Principal component analysis

To account for breed population structure, a principal component analysis (PCA) was executed, and the resultant breed dimensions used as covariates. To run the PCA, the phenotype and covariate PED/MAP files from PLINK were converted to binary files using the command “*plink --noweb --dog --file PBQgenotype --out PBQgenotype --geno 0.1 --mind 0.05 --maf 0.01 --hwe 0.001 --make-bed*”.

Subsequent steps were run using Genome-Wide Complex Trait Analysis (GCTA) software (Yang *et al.*, 2011). The binary files were used to create a genomic relationship matrix (GRM) for autosomal chromosomes using the command: *gcta64 --bfile PBQgenotype --autosome-num 38 --autosome --make-grm --out PBQgenotype*. Where “--bfile” indicates the input of binary files, “--autosome-num 38” indicates the 38 autosomal chromosome pairs in dogs, “--autosome” is used to base the relationship matrix on the entire autosome rather than a single chromosome and excluding the sex chromosomes, and “--make-grm” is used to specify the process. The GRM was then used, again in GCTA, to perform a PCA using the command “*gcta64 --grm PBQgenotype --autosome-num 38 --autosome --pca 20 --out PBQgenotype*”. The PCA resulted in the extraction of 20 breed dimensions (BDs). The first and second breed dimensions were plotted as a scatter plot to visualise the structure of the population. Breed dimensions 1 and 2 were used in covariate files to account for breed within the model.

6.3.6.3 Population stratification

Population stratification was visualised with the aid of a multidimensional scaling (MDS) plot. QQ and Manhattan plots were created using the *qqman* package in R (version 2023.06.1). QQ plots were created using the *qqnorm()* function. Manhattan plots were created using the *manhattan()* function.

6.3.6.4 *Restricted Estimate of Maximum Likelihood heritability analysis*

A Restricted Estimate of Maximum Likelihood (REML) model was run for genotype heritability analyses using GCTA. To select the most appropriate heritability model different combinations of environmental factors were reviewed, based on significant findings from Chapter 3. The most basic model was run for consistency with the pedigree heritability models for comparisons as covariates had little influence on the data set due to the small sample size. The final model included sex and breed dimensions 1 and 2. Due to the small data size, it was not possible to add random factors such a puppy raiser experience and presence of children, which were shown as significant factors in Chapter 3.

A separate model was run for each r-PWQ2 component at 8M and 12M using the command *“gcta64 --reml --grm PBQgenotype --pheno 12Mdata.txt --mpheno 1 --qcovar 12Mcovar.txt --out PBQgenotype”*.

Where “reml” indicates that a Restricted Estimate of Maximum Likelihood model and “grm” a genetic relationship matrix was used. The “mpheno” indicates the phenotype number used (which is component 1 in the example above) and “qcovar” indicates the covariate file used (in the example above this is the twelve-month covariate file containing sex and breed dimensions 1 and 2). Genomic-based heritability estimates are reported as $\pm$ standard error.

6.3.6.5 *Genome-wide association analysis (GWAS)*

For the GWAS, the GCTA software was used to run a mixed linear model analysis (MLMA). A separate command was run for each r-PWQ2 component for 8M and 12M data. The command *“gcta64 --mlma --bfile PBQgenotype --autosome-num 38 --autosome --grm PBQgenotype --pheno 12Mdata1.txt --mpheno 1 --qcovar 12Mcovar.txt --out PBQgenotype”* was run. Where “mlma” indicates that a mixed linear model analysis was run and “autosome num 38” that the dog genome with 38 autosomes and “grm” a genetic relationship matrix was used. The “mpheno” indicates the phenotype number used (which is component 1 in

the example above) and “qcovar” indicates the covariate file used (in the example above this is the 12M covariate file containing sex and breed dimensions 1 and 2).

The top 20 significant SNPs and base pair position information was extracted for behaviour components that showed genomic-based heritabilities above 0.1. The ability to run bivariate models to review genetic correlations between each component was checked, however the data set was too small to successfully run analyses.

Ensembl genome browser (<https://www.ensembl.org/index.html>) was used to identify genes (Ensemble Genes 112) of interest when compared to reference dog genome (ROS_CanFam_1.0). First the chromosome and region of interest was determined and entered into the ‘BioMart’ tool and a multi species comparison made to orthologous human genes. The chromosome/scaffold name, gene start (bp), gene end (bp) and gene name were extracted as Excel files to provide a list of genes of interest.

A list of genes for each of six components were then entered into the Database for Annotated, Visualisation and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/tools.jsp>), a bioinformatics knowledgebase used to translate findings to biological meaning (Dennis *et al.*, 2003). Lists of genes were compared findings from human (*Homo sapiens*) research. The functional annotation tool was used to search areas of interest, which included disease (GAD_DISEASE, OMIM_DISEASE, UP_KW_DISEASE), functional annotations (UP_KW_BIOLOGICAL_PROCESS, UP_KW_CELLULAR_COMPONENT, UP_KW_MOLECULAR_FUNCTION, UP_KW_PTM, UP_SEQ_FEATURE), gene ontology (GOTERM_BP_DIRECT, GOTERM_CC_DIRECT, GOTERM_MF_DIRECT) and tissue expression (HPA_NORMAL_TISSUE, HPA_RNA_TISSUE, UP_TISSUE). Outputs were analysed and clusters with the highest environment scores for relevant biological pathways, disease associations or tissue expression

were noted. The enrichment score was reported, which is a measure of how frequently specific terms are represented in a cluster compared with other clusters found.

Following a review of QQ and Manhattan Plots, the components with suggested significance were explored further. The gene closest to the significant chromosome and base pair position was extracted and links to behavioural phenotypes published in the literature were explored using PubMed (<https://pubmed.ncbi.nlm.nih.gov/>) and GWAS catalog (<https://www.ebi.ac.uk/gwas/home>).

6.4 Results

6.4.1 r-PWQ2 data

For the 96 dogs selected with complete 12M data, 81 dogs also had 8M r-PWQ2 scores, so genotype analysis was undertaken using both phenotype data sets. Dogs in the sample represented a spread of behaviour scores at both 12M (Figure 6.2) and 8M (Figure 6.3).

At 12M the average component score for the 96 dogs ranged from 4.26 (± 5.93 S.D.) for C1 (Separation) to 84.77 (± 15.92 S.D.) for C8 (Energy). The minimum score was below 1.00 for nine of the 10 components and the maximum score was over 80.00 for six components. The largest range was for C7 (Anticipatory Excitement) with a minimum score of 0.00 and a maximum of 100.00.

At 8M the average component score for the 81 dogs ranged from 6.49 (± 8.92 S.D.) for C3 (Separation) to 84.45 (± 15.35 S.D.) for C7 (Energy). The minimum score was 0.00 for six components and the maximum score was over 85.00 for five components. The largest range was for component 5 (Anticipatory Excitement) with a minimum score of 0.00 and a maximum of 99.50.

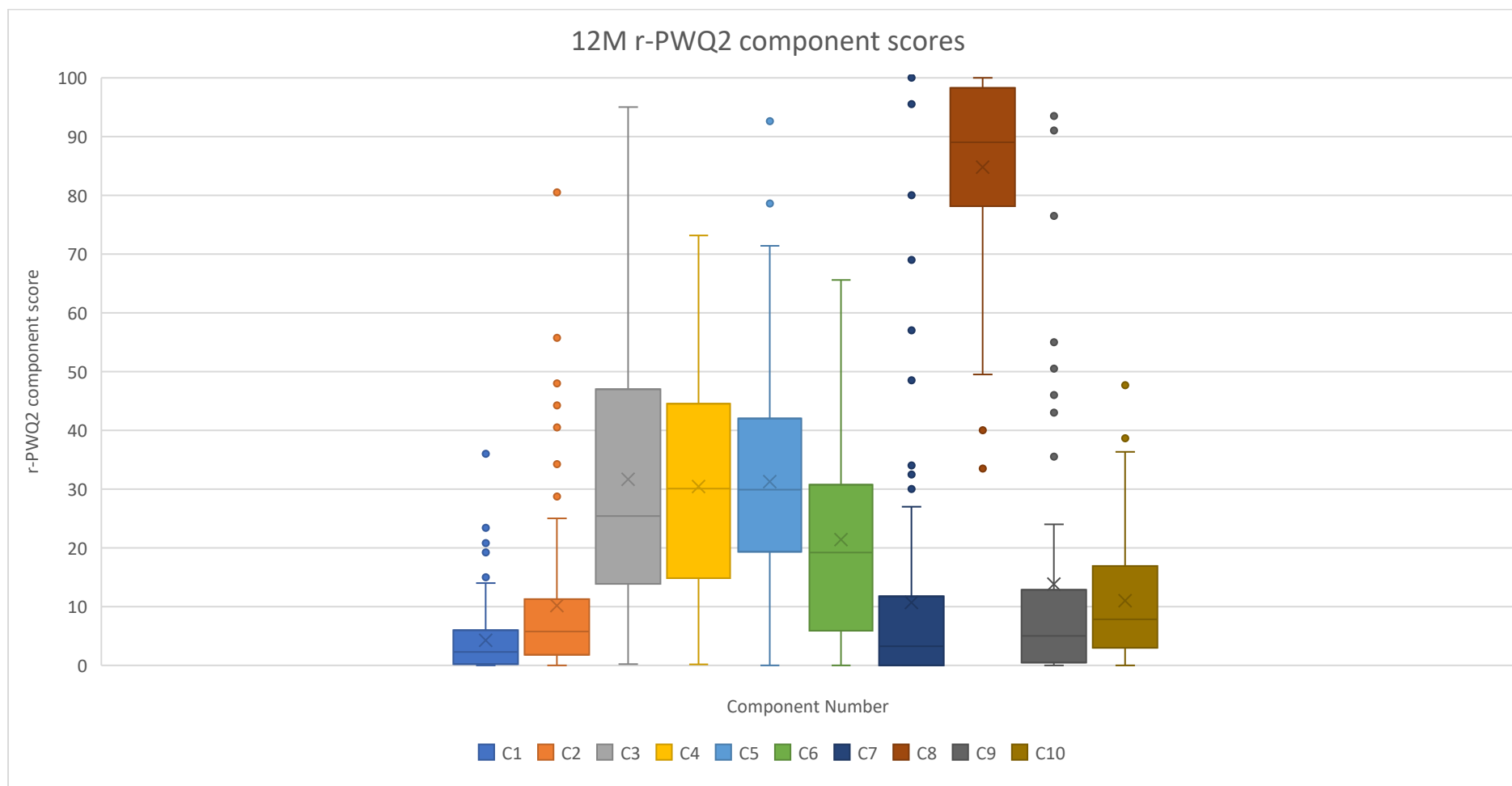


Figure 6.1 12M r-PWQ2 component score distribution for dogs selected as part of genotype analyses using a box and whisker plot, indicating outliers and range.

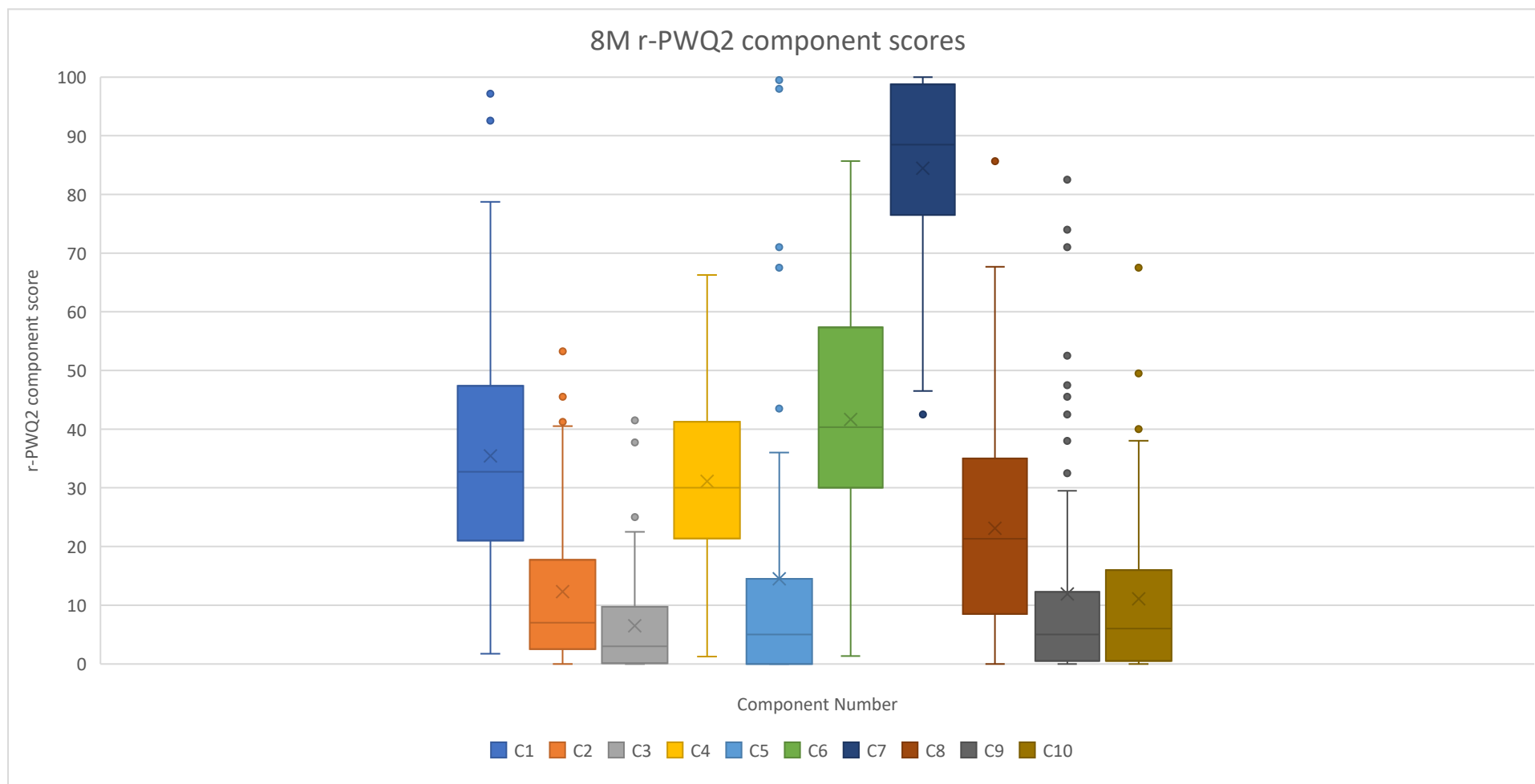


Figure 6.2 8M r-PWQ2 component scores distribution for dogs selected as part of genotype analyses using a box and whisker plot, indicating outliers and range.

6.4.2 Sample quality control checks

The first step determined sex for genotyped individuals. Eighty-two dogs had sex correctly imputed from X chromosome SNPs, but for 14 dogs (10 males/4 females) F was outside the thresholds needed to impute sex. Figure 6.4 represents the spread of F across individuals in the data set, with those above the red line indicating the dog to be male and those below the red line that the dog was female. The average F (heterozygosity) was 0.34 (± 0.04 S.E., 95% CI 0.25 to 0.44) across the entire sample. The X chromosome inbreeding coefficients for males ranged from 0.75 to 0.89 (mean 0.82, ± 0.00 S.E., 95%CI 0.81 to 0.82) and for females from -0.54 to 0.55 (mean -0.07, ± 0.03 S.E., 95% CI -0.12 to -0.01). There were three dogs with a X chromosome inbreeding coefficient above 0.4 of which all were female Labradors. The inbreeding coefficients of the animals were checked and results were within normal acceptable limits (0.04-0.06).

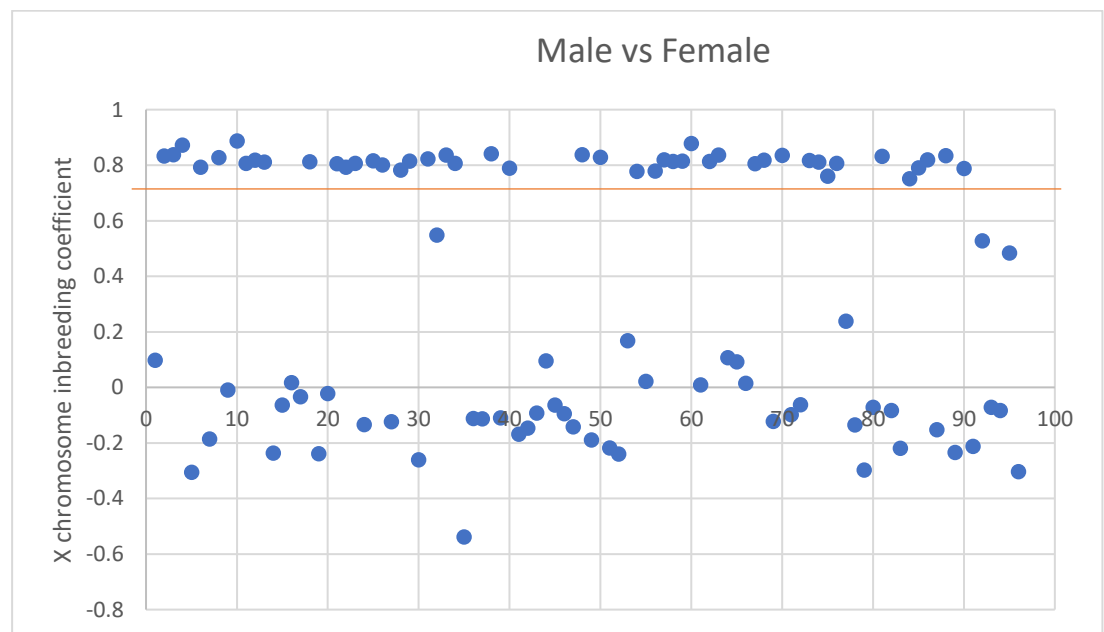


Figure 6.3 Heterozygosity estimates for genotypes samples.

The second step determined SNPs missing from the 96 samples. The number (proportion) of missing SNPs ranged from 2795 (0.012) for

individual 21 to 6342 (0.028) for individual 90. There were 220,853 SNPs available for all 96 samples.

Step 3 identified the observed rate of heterozygosity per individual. Observed number of homozygous genotypes ranged from 101,608 to 119,369 and the number of non-missing genotypes per individual ranged from 163,280 to 165,744. Individual observed heterozygosity ranged from 0.28 to 0.38, with a mean of 0.35 (± 0.03). Heterozygosity for females was 0.34 (± 0.04) and for males was 0.35 (± 0.03). Heterozygosity for Labradors was 0.30 (± 0.01) and for Labradors crossed with Golden Retrievers (Labrador x Golden Retriever or Golden Retriever x Labrador) was 0.37 (± 0.00). Observed heterozygosity for all individuals were less than 3 standard deviations from the mean and therefore there were no outliers to be excluded from further analysis.

Step 4 pruned the SNPs so that for correlated SNPs (those in high linkage disequilibrium with each other) only one was retained in the analysis set. There were 121,150 SNPs pruned due to being in linkage disequilibrium (LD) with neighbouring SNPs, with 97,105 SNPs remaining.

Step 5 identified any duplicated or related individuals in the sample. The pairwise identity-by-state (IBS) distances between individuals, indicative of relatedness, were evaluated and IBS ranged from 0 to 0.44, concluding that there were no duplicated individuals in the sample. The average relationship between individuals was 0.11 (± 0.08).

A plot of individuals based on IBS sharing was created to visualise any population stratification (Figure 6.5). A clear distinction between pure Labradors (green dots) and crossbreeds (blue dots) was noted so breed dimension outputs were included in further analyses.

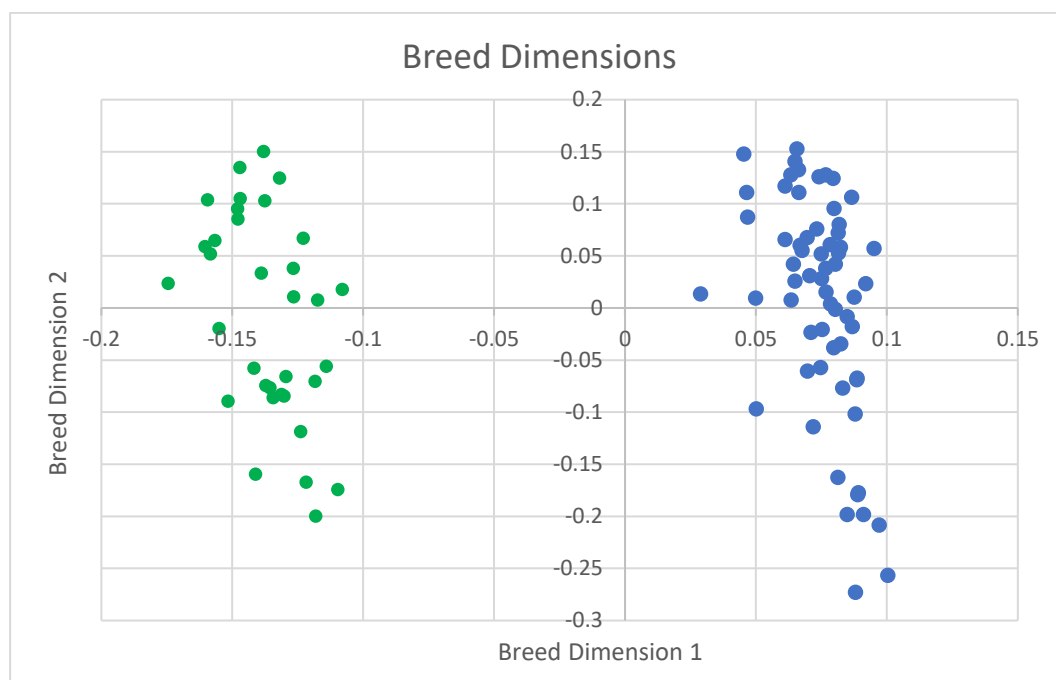


Figure 6.4 Breed dimensions outputs for pure and crossbreeds. Blue dots indicate crossbreed dogs and green dots purebred dogs in the sample.

Step 6 identified markers with an excessive missing data. The results enabled filter thresholds to be determined for missing rates per SNP and per individual and for the Hardy Weinberg Equilibrium (HWE). When SNP missing rate was set to 90% (individuals with at least 90% of the SNPs successfully genotyped) 7465 SNPs failed the missingness test. When excluding SNPs that were not present in over 95% of individuals, 2515 SNPs failed to reach the required threshold. A further 4804 markers were excluded based on HWE test ($P < 0.001$). After frequency and genotyping pruning, there were 211,423 SNPs left in the analyses. Following the exclusion of SNPs with a minor allele frequency below 1%, 160,598 SNPs remained for GWAS analyses.

Once the QC controls had been completed, the possibility of undertaking separate GWAS for purebreds ($n=34$) and crossbreeds ($n=62$) was explored. The sample numbers were too small for accurate comparisons to be made, so all GWAS were completed on the whole sample.

6.4.3 12M REML heritability analysis

The 12M REML heritability analyses were undertaken on 96 dogs. Genetic variance (V_g) was greatest for C7 (Anticipatory Excitement; 184.34) and was 0.00 for six components. Environmental variance (V_e) was higher than genetic variance, with the greatest environmental variance found for C3 (541.95; Impulsivity) and 9 (487.89; Body Sensitivity). Phenotype variance (V_p) was also greater than genetic variance and ranged from 36.35 (C1; Separation) to 541.95 (Impulsivity; C3). Genomic-based heritability estimates (h^2) for 12M components ranged from 0.00 to 0.44 (Table 6.1), with four showing a heritability estimate greater than zero. P-values for all components were above 0.1.

Table 6.1 12M Genomic-based heritability estimates for 12M r-PWQ2 components.

Component	h^2	$\pm$ S.E.	V_g	V_e	V_p	P-value
C1	0.04	0.42	1.56	34.79	36.35	0.46
C2	0.00	0.38	0.00	159.77	159.77	0.50
C3	0.00	0.41	0.00	541.95	541.95	0.50
C4	0.00	0.47	0.00	285.29	285.29	0.50
C5	0.00	0.43	0.00	296.72	296.72	0.50
C6	0.43	0.39	143.79	189.73	333.52	0.11
C7	0.44	0.38	184.34	232.54	416.88	0.11
C8	0.19	0.41	50.63	211.69	262.32	0.31
C9	0.00	0.44	0.00	487.89	487.89	0.50
C10	0.00	0.42	0.00	101.29	101.29	0.50

Heritability estimates (h^2) and their standard errors (S.E), the genetic variance (V_g), the environmental variance (V_e), the phenotypic variance (V_p) and the P-value are all shown to two decimal places.

Comparisons of the genomic-based heritability estimates to those calculated using pedigree estimates (Chapter 3), highlighted some similarities (Table 6.2). The highest pedigree heritability estimates were found to be no higher than 0.08 (C2, C6, C9) whilst genomic-based estimates were found to be as high as 0.44 (C7). Only three genomic-based heritability estimates were greater than 0.10 (C6, C7, C8), with all three demonstrating pedigree-based heritability estimates of 0.07 or

0.08. Two components (C2 and C9) had pedigree-based heritability estimates of 0.08 and contrasting genomic-based heritability estimates of 0.00. No heritability estimates (for both genomic-based and pedigree-based) were significant.

Table 6.2 12M Genomic-based and pedigree-based heritability estimates for 12M r-PWQ2 components.

Component	r-PWQ2 component	Genomic-based h^2	$\pm$ S.E.	Pedigree-based h^2	$\pm$ S.E
C1	Separation	0.04	0.42	0.06	0.05
C2	Anxiety	0.00	0.38	0.08	0.05
C3	Impulsivity	0.00	0.41	0.00	0.00
C4	Focus	0.00	0.47	0.00	0.00
C5	Attachment & Attention Seeking	0.00	0.43	0.03	0.04
C6	General Excitement	0.43	0.39	0.08	0.06
C7	Anticipatory Excitement	0.44	0.38	0.07	0.05
C8	Energy	0.19	0.41	0.07	0.05
C9	Body Sensitivity	0.00	0.44	0.08	0.05
C10	Human Focus	0.00	0.42	0.04	0.04

Heritability estimates (h^2) and their standard errors (S.E.) are shown to two decimal places.

6.4.4 12M GWAS

There were three components at 12M with a genomic-based heritability estimate above 0.1: C6 (General Excitement, $h^2=0.43$), C7 (Anticipatory excitement, $h^2=0.44$) and C8 (Energy, $h^2=0.19$). These three components were explored in further detail to determine significant regions of interest associated with the behavioural measures.

6.4.4.1 *Component 6 - General Excitement*

Component 6 (C6) had a genomic-based heritability estimate of 0.43 (± 0.39) and a pedigree heritability estimate of 0.08 (± 0.06) at 12M. The

top 20 significant SNPs associated with Component 6 were found on a range of chromosomes (Table 6.3). SNPs were located on Chromosomes 2, 3, 6, 8, 12, 13, 15, 17, 18, 21, 22, 24, 29 and 36. A cluster of three SNPs was found on Chromosome 17 between the region of 35.9-53.4 Mb (megabases).

Table 6.3 Top 20 significant SNPs for C6 at 12M. Chromosome number, SNP, base pair position and P-value are reported.

Chromosome	SNP	Region (bp position)	P-value
2	BICF2S2317492	45780662	0.00050
2	BICF2S2317574	67725599	0.00030
3	BICF2G630108179	36222317	0.00046
6	BICF2P1418220	39022244	0.00023
8	TIGRP2P110568_rs8922393	24594252	0.00043
8	BICF2P1250542	71515413	0.00017
12	BICF2S23537932	19860319	0.00005
12	TIGRP2P169773_rs9182056	58308652	0.00027
13	BICF2P405455	47359636	0.00043
15	chr15_37518008	37518008	0.00024
17	BICF2P1165732	35992400	0.00045
17	BICF2G630202403	48817677	0.00001
17	BICF2P195088	53464522	0.00001
18	BICF2P1288679	30211094	0.00011
21	BICF2G630658559	6062929	0.00033
22	BICF2G630325855	25428554	0.00035
24	BICF2P1128018	11687550	0.00019
29	BICF2G630622521	11592469	0.00043
29	BICF2G630622529	11602139	0.00046
36	BICF2P904180	22926993	0.00045

The QQ plot (Figure 6.6) showed a deviation below the normal distribution line, and no SNPs were noted to have an effect of suggestive significance in the Manhattan plot (Figure 6.7). Results suggest that there were no SNPs with significant effect on C6 at 12M. An in-depth exploratory analysis of the 20 significant SNPs was therefore not undertaken.

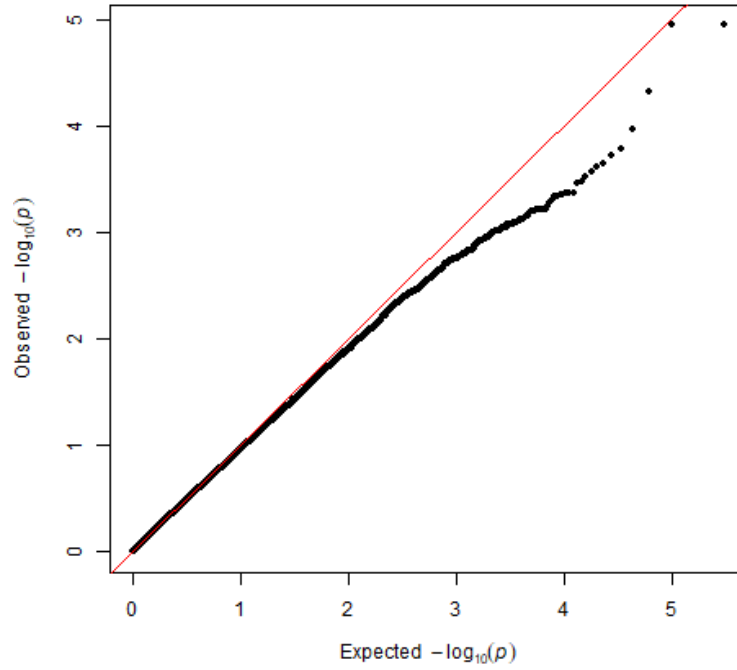


Figure 6.5 QQ plot for C6 at 12M.

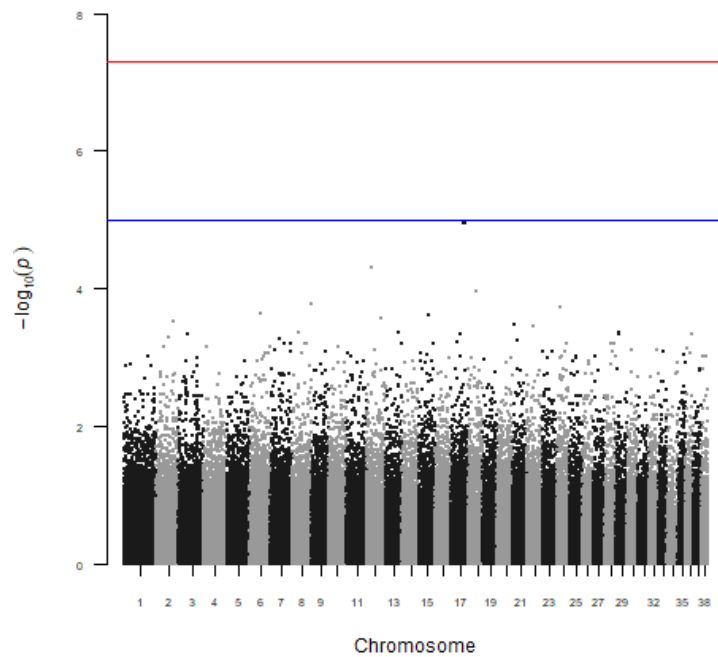


Figure 6.6 Manhattan plot for C6 at 12M.

6.4.4.2 Component 7 - Anticipatory Excitement

Component 7 (C7) had a genomic-based heritability estimate of 0.44 (± 0.38) and a pedigree heritability estimate of 0.07 (± 0.05) at 12M. The top 20 significant SNPs associated with Component 7 were found on a

range of chromosomes (Table 6.4). SNPs were located on Chromosomes 2, 4, 7, 8, 9, 10, 13, 18, 19, 20, 22, 24, 25, 29, 34 and, 36. A cluster of five SNPs was found on Chromosome 29 between the region of 11.5-19.4Mb and around 36.4Mb.

Table 6.4 Top 20 significant SNPs for C7at 12M. Chromosome number, SNP, base pair position and P-value are reported

Chromosome	SNP	Region (bp position)	P-value
2	BICF2S23760457	37209830	0.00001
4	BICF2G630167772	36746011	0.00001
7	TIGRP2P103126_rs9156626	72038443	0.00000
8	chr8_47570702	47570702	0.00002
9	BICF2P378226	13616096	0.00004
10	chr10_35753478	35753478	0.00001
13	chr13_47717309	47717309	0.00001
18	BICF2S2443140	4899169	0.00001
19	BICF2P1177994	42024137	0.00001
20	BICF2S23435805	6739738	0.00001
22	chr22_30607147	30607147	0.00002
24	BICF2P303373	42438972	0.00004
25	BICF2P609947	38586141	0.00001
29	BICF2P341807	11558277	0.00000
29	BICF2P344431	15783184	0.00001
29	BICF2P344397	15834948	0.00001
29	BICF2P344367	19467034	0.00002
29	BICF2P332032	36475317	0.00002
34	TIGRP2P401046_rs8532675	39420537	0.00001
36	BICF2P860086	9107151	0.00002

The QQ plot (Figure 6.8) demonstrated significant association of SNPs, due to a deviation above the normal distribution line. Similarly, individual SNPs with significance of association exceeding the line denoting suggestive significance in the Manhattan plot (Figure 6.9) can be observed. An in-depth exploratory analysis of the 20 significant SNPs was therefore undertaken.

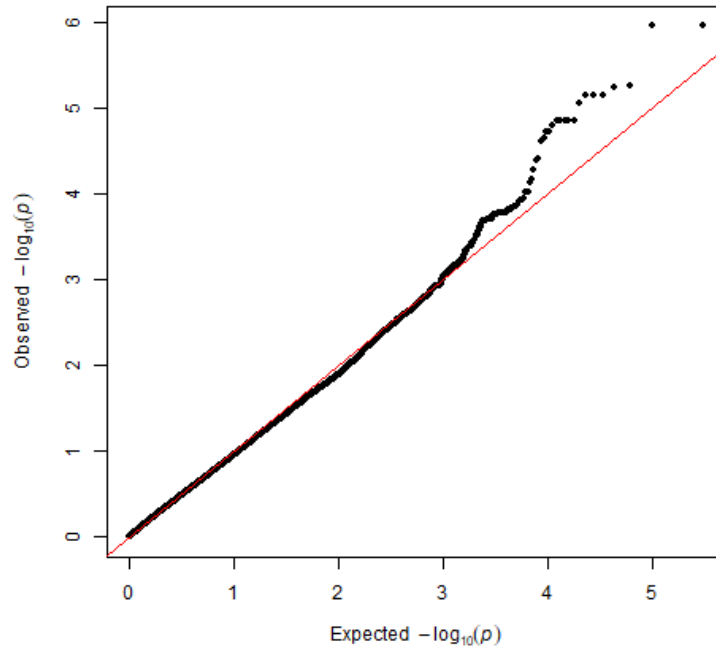


Figure 6.7 QQ plot for C7 at 12M.

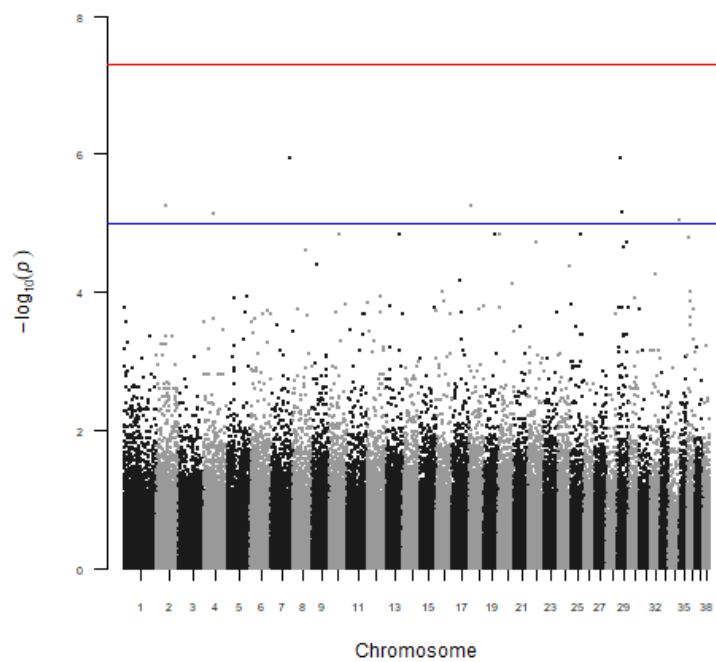


Figure 6.8 Manhattan plot for C7 at 12M.

After all genes identified in the regions of interest were submitted to Database for Annotation, Visualization, and Integrated Discovery (DAVID), the highest scoring functional annotations/ terms/ diseases of interest were:

- Calcium-dependent protein binding (enrichment score 3.8)
- Cholesterol biosynthesis (enrichment score 1.9)
- Regulation of dopamine secretion (enrichment score 0.7)
- Alzheimer's disease (enrichment score 0.65)

Genes in the associated regions have been linked to a range of behavioural conditions in humans for C7 at 12M. These include Schizophrenia, Alzheimer's disease, Autism, alcoholism, mood disorders and eating disorders (e.g. anorexia nervosa). A summary of genes in the genome regions of interest are listed in Table 6.5.

Table 6.5 Associated behavioural phenotypes for C7 at 12M. The chromosome number, base pair position (Mb) of SNP(s) and the nearest gene are reported for the 20 SNPS with lowest P-values.

Chromosome	Mb of SNPs with lowest P-values	Nearest gene(s)	Associated behavioural phenotype (in humans unless species stated)
2	37.20	PCDH1	Brain development (Kim <i>et al.</i> , 2007), emotional and social behaviour (Krishna <i>et al.</i> , 2011), cognitive function (Lee <i>et al.</i> , 2018), Dyslexia (Naskar <i>et al.</i> , 2018), Schizophrenia (Shao <i>et al.</i> , 2019), Autism spectrum disorder (Zheng <i>et al.</i> , 2020), Bipolar disorder (Fjukstad <i>et al.</i> , 2021), memory performance (Eissman <i>et al.</i> , 2024)
4	36.74	UIMC1	Neurodevelopmental disorders (Wang <i>et al.</i> , 2020)
7	72.03	EPB41L3	Axon organisation and myelination (Einheber <i>et al.</i> , 2013), social communication impairment (St Pourcain <i>et al.</i> , 2014), tobacco use disorder (Sherman <i>et al.</i> , 2022), Alzheimer's disease (Sun <i>et al.</i> , 2024), developmental disorder (Werren <i>et al.</i> , 2024)
8	47.57	LIN52	Educational attainment (Lee <i>et al.</i> , 2018), stress response (González-Rangel and Navarro, 2021) cognitive function (Homann <i>et al.</i> , 2022), tobacco use disorder (Sherman <i>et al.</i> , 2022)
9	13.61	ERN1	Alzheimer's disease (Montibeller and de Belleruche, 2018; He <i>et al.</i> , 2021a), major depressive disorder (Giannakopoulou <i>et al.</i> , 2021)
10	35.75	EDAR, SH3RF3	Alzheimer's disease (Naj <i>et al.</i> , 2010), Schizophrenia (Goes <i>et al.</i> , 2015; Wu <i>et al.</i> , 2020), cognitive impairment (Jia <i>et al.</i> , 2017), intelligence (Davies <i>et al.</i> , 2018), dyslexia (Gialluisi <i>et al.</i> , 2019), education attainment (Okbay <i>et al.</i> , 2022), tobacco use disorder (Sherman <i>et al.</i> , 2022), cognitive function (Acharya <i>et al.</i> , 2023)
13	47.72	KIT	Bipolar disorder (He <i>et al.</i> , 2019), obsessive compulsive disorder and anorexia nervosa (Yilmaz <i>et al.</i> , 2020)
18	4.89	No known genes	-

19	42.02	HNMT	Alcoholism (Oroszi <i>et al.</i> , 2005), ADHD (Stevenson <i>et al.</i> , 2010), aggression and sleep-wake cycle (Naganuma <i>et al.</i> , 2017), autism spectrum disorder (Wright <i>et al.</i> , 2017), smoking initiation (Liu <i>et al.</i> , 2019; Saunders <i>et al.</i> , 2022), intellectual disability (Verhoeven <i>et al.</i> , 2020; Taşkıran <i>et al.</i> , 2021), brain function (Otsuka <i>et al.</i> , 2022)
20	6.73	TAMM41, VGLL4	Anorexia nervosa (Wang <i>et al.</i> , 2011; Clarke <i>et al.</i> , 2014), neuroticism (Kim <i>et al.</i> , 2017), depressive behaviours (Guo <i>et al.</i> , 2024)
22	30.60	KCTD12	Aggression and mating behaviour (Williams <i>et al.</i> , 2014), mood disorders (Cathomas <i>et al.</i> , 2015), Autism (Leblond <i>et al.</i> , 2019), bipolar disorder (Teng <i>et al.</i> , 2019), stress (Deng <i>et al.</i> , 2021), memory performance (Kang <i>et al.</i> , 2023)
24	42.43	RTF2, GCNT7, U6	Drug and alcohol abuse (Navarro, 2017), eating behaviours (Micioni Di Bonaventura <i>et al.</i> , 2020), anorexia (Sweeney <i>et al.</i> , 2021), Alzheimer's disease (Adewuyi <i>et al.</i> , 2022; Gouveia <i>et al.</i> , 2022), suicidal behaviours (Kimbrel <i>et al.</i> , 2023). The region between 41.97 – 42.33Mb contains MC3R which has been associated with anxiety (Cho <i>et al.</i> , 2023)
25	38.58	DOCK10, NYAP2	Autism (Guo <i>et al.</i> , 2017), smoking behaviour (Karlsson Linnér <i>et al.</i> , 2019), ADHD and substance abuse (Karlsson Linnér <i>et al.</i> , 2021), mood disorders (Le-Niculescu <i>et al.</i> , 2021), Dementia (Fongang <i>et al.</i> , 2024)
29	11.55-36.47	~105 genes in region	See below
34	39.42	NAALADL2	Autism (Kuo <i>et al.</i> , 2015), depression (Lin <i>et al.</i> , 2018), cognitive impairment (Lutz <i>et al.</i> , 2019), Schizophrenia (Trubetskoy <i>et al.</i> , 2022), neurodevelopmental disorders (Schrauwen <i>et al.</i> , 2024)
36	9.10	FIGN	Major depressive disorder (Aragam <i>et al.</i> , 2011), risk taking behaviour (Karlsson Linnér <i>et al.</i> , 2019), neuroticism and cognitive function measurements (Hindley <i>et al.</i> , 2023)memory performance (Eissman <i>et al.</i> , 2024)

Running 105 genes from the region on Chromosome 29 (11.55-36.47Mb) through DAVID, to search for KEGG and REACTOME pathways and UP Tissue indicated that pathways with the highest enrichment score (3.03) were metabolic pathways (eight genes), reversible hydration of carbon dioxide (four genes) and nitrogen metabolism (four genes). The Chromosome 29 regions contained several genes related to gene expression (transcription; nine genes), disease (eight genes), adaptive immune systems (six genes).

6.4.4.3 Component 8 – Energy

Component 8 (C8) had a genomic-based heritability estimate of 0.19 (± 0.41) and a pedigree heritability estimate of 0.07 (± 0.05) at 12M. The top 20 significant SNPs associated with C8 were found on a range of chromosomes (Table 6.6). SNPs were located on Chromosomes 3, 4, 5, 6, 12, 14, 21, 25, 28, 33 and, 35. A cluster of three SNPs were found on Chromosome 3 at approximately 2.3, 45.4 and 78.9Mb. Four SNPs were found on Chromosome 4 at 2.0Mb, 41.4Mb, 49.9Mb and 61.5Mb. A cluster of three SNPs were found for Chromosome 5 between 7.8-11.4Mb and another at around 80.8Mb.

Table 6.6 Top 20 significant SNPs for C8 at 12M. Chromosome number, SNP, base pair position and P-value are reported.

Chromosome	SNP	Region (bp position)	P-value
3	BICF2P640419	2383010	0.00002
3	BICF2P646035	45453985	0.00009
3	BICF2G630353474	78960101	0.00011
4	BICF2S2354158	2075582	0.00021
4	BICF2S23542486	41459104	0.00021
4	BICF2P855052	49928668	0.00011
4	BICF2P897013	61558542	0.00011
5	BICF2S23544923	7895807	0.00002
5	BICF2S23543795	10022182	0.00019
5	BICF2S235436	11439908	0.00019
5	BICF2P1369962	80862588	0.00007
6	chr6_19496330	19496330	0.00011
12	BICF2P1375381	15736133	0.00029

14	BICF2S2391388	18840271	0.00023
21	chr21_23474568	23474568	0.00017
25	BICF2P99615	3672985	0.00011
25	BICF2S23048982	43683673	0.00011
28	BICF2P823273	23428331	0.00001
33	chr33_17404231	17404231	0.00023
35	BICF2S23256178	6538999	0.00026

The QQ plot (Figure 6.10) showed a deviation below the normal distribution line and no SNP effects exceeded the line denoting suggestive significance on the Manhattan plot (Figure 6.11). Results suggest that there were no SNPs with significant effect on C8 at 12M. Therefore, an in-depth exploratory analysis of the 20 significant SNPs was not undertaken.

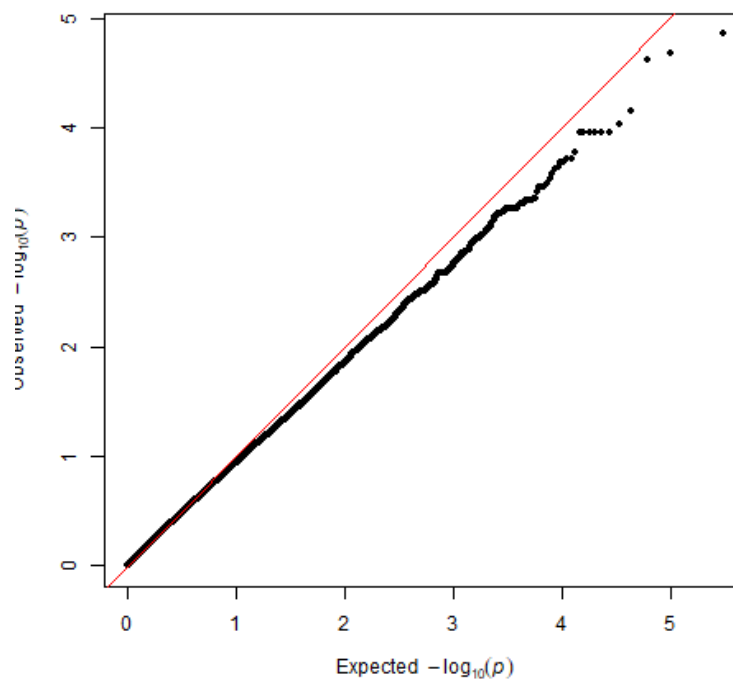


Figure 6.9 QQ plot for C8 at 12M.

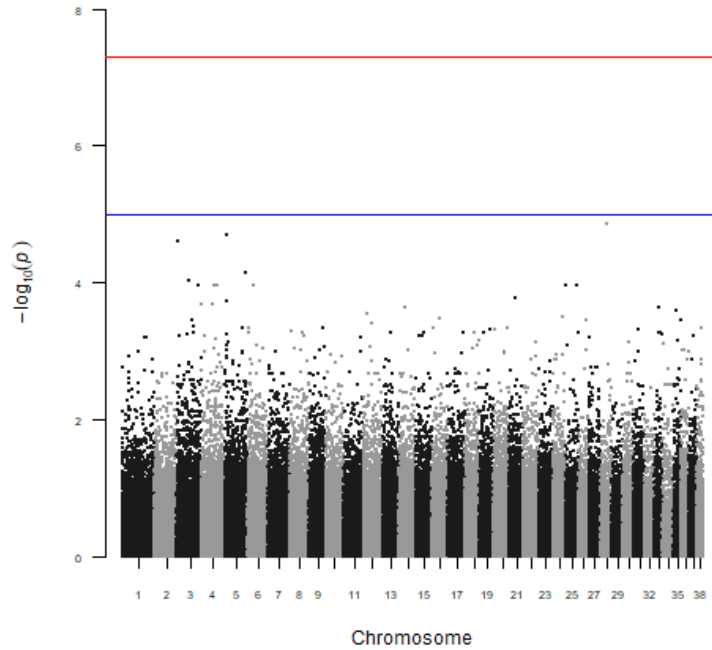


Figure 6.10 Manhattan plot for C8 at 12M.

6.4.5 8M REML heritability analysis

The 8M REML heritability analyses were undertaken on 81 dogs. Genetic variance (V_g) was greatest for C8 (Focus; 293.93) and was 0.00 for seven components. Environmental variance (V_e) was higher than genetic variance, with the greatest environmental variance found for C1 (Impulsivity; 421.99). Phenotype variance (V_p) was higher than both genetic and environmental variation, ranging from 86.12 (C3; Separation) to 644.77 (C5; Anticipatory Excitement). Genomic-based heritability estimates for 8M components ranged from 0.00 to 0.91 (Table 6.7), with three showing a heritability estimate greater than zero. P-values for all components were above 0.05.

Table 6.7 8M Genomic-based heritability estimates for 8M r-PWQ2 components.

Component	h^2	$\pm$ S.E.	V_g	V_e	V_p	P-value
C1	0.00	0.46	0.00	421.99	421.99	0.50
C2	0.00	0.41	0.00	163.32	163.32	0.50
C3	0.62	0.44	53.23	32.90	86.13	0.10
C4	0.00	0.50	0.00	216.30	216.30	0.50

C5	0.35	0.49	223.74	421.03	644.77	0.25
C6	0.00	0.45	0.00	405.16	405.16	0.50
C7	0.00	0.46	0.00	209.94	209.94	0.50
C8	0.91	0.46	293.93	30.12	324.05	0.08
C9	0.00	0.50	0.00	269.79	269.79	0.50
C10	0.00	0.52	0.00	167.80	167.80	0.50

Heritability estimates (h^2) and their standard errors (S.E), the genetic variance (V_g), the environmental variance (V_e), the phenotype variance (V_p) and the P-value are all shown to two decimal places.

Comparisons of the genomic-based heritability estimates to those calculated using pedigree estimates (Chapter 3), highlighted some similarities (Table 6.8). The highest pedigree heritability estimates were found to be no higher than 0.09 (C5) whilst genomic-based estimates were found to be as high as 0.91 (C8). Three genomic-based heritability estimates were greater than 0.30 (C3, C5, C8), with pedigree-based heritability estimates ranging from 0.03 to 0.09. Seven genomic-based heritability estimates were found to be 0.0, indicating no correlations between the phenotype and SNPs studied. C1 had a pedigree-based heritability estimate of 0.08 and contrasting genomic-based heritability estimate of 0.00. Two pedigree-based heritability estimates were significant (C1 and C5), whilst no genomic-based heritability estimates were significant.

Table 6.8 8M Genomic-based and pedigree-based heritability estimates for 8M r-PWQ2 components.

Component	r-PWQ2 component	Genomic-based h^2	$\pm$ S.E.	Pedigree-based h^2	$\pm$ S.E.
C1	Impulsivity	0.00	0.46	0.08	0.04
C2	Anxiety	0.00	0.41	0.02	0.03
C3	Separation	0.62	0.44	0.05	0.04
C4	Distraction	0.00	0.50	0.02	0.03
C5	Anticipatory Excitement	0.35	0.49	0.09	0.04
C6	Attachment	0.00	0.45	0.05	0.04
C7	Energy	0.00	0.46	0.04	0.03
C8	Focus	0.91	0.46	0.03	0.03
C9	Body Sensitivity	0.00	0.50	0.03	0.03

C10	Attention Seeking	0.00	0.52	0.00	0.03

Heritability estimates (h^2) and their standard errors (S.E.) are shown to two decimal places.

6.4.6 8M GWAS

There were three components at 8M with a genomic-based heritability estimate above 0.1: C3 (Separation; $h^2=0.62$), C5 (Anticipatory excitement, $h^2=0.35$) and C8 (Focus, $h^2=0.91$). These three components have been explored in further detail to determine significant regions of interest associated with the behavioural measures.

6.4.6.1 *Component 3 – Separation*

Component 3 (C3) had a genomic-based heritability estimate of 0.62 (± 0.44) and a pedigree heritability estimate of 0.05 (± 0.04) at 8M. The top 20 significant SNPs associated with C3 were found for a range of chromosomes (Table 6.9). SNPs were located on Chromosomes 2, 3, 5, 9, 11, 13, 15, 18, 19, 21, 25, 30 and 32. Two clusters of SNPs were detectable on Chromosome 5; four SNPs between 8.4-20.0Mb and four between 53.9-75.0Mb.

Table 6.9 Top 20 significant SNPs for C3 at 8M. Chromosome number, SNP, base pair position and P-value are reported.

Chromosome	SNP	Region (bp position)	P-value
2	BICF2P1068421	79491895	0.0000209
3	BICF2S23210420	46220253	0.0000148
5	BICF2S23221365	8457618	0.0000005
5	BICF2S23214056	13364926	0.0000005
5	BICF2S22958286	17439850	0.0000423
5	BICF2S23219334	20014985	0.0000005
5	BICF2S22960475	53955431	0.0000423
5	BICF2S23219015	63325407	0.0000005
5	BICF2S23221966	74672586	0.0000005
5	BICF2S23220618	75063484	0.0000005
9	BICF2G630839920	39885813	0.0000002
11	BICF2P929963	28648902	0.0000001

13	BICF2P1216158	46013356	0.0000237
15	BICF2P356564	27421154	0.0000004
18	BICF2P701996	55295783	0.0000227
19	BICF2S23160231	45011136	0.0000227
21	BICF2P352732	47638780	0.0000001
25	BICF2P617416	29621450	0.0000002
30	BICF2P549092	30698652	0.0000006
32	BICF2P216858	2231037	0.0000227

The QQ plot (Figure 6.12) showed significant SNPs, seen by a deviation above the normal distribution line. Similarly, eight individual SNPs demonstrated effects exceeding the line, denoting suggestive significance in the Manhattan plot (Figure 6.13). An in-depth exploratory analysis of the top 20 significant SNPs was undertaken.

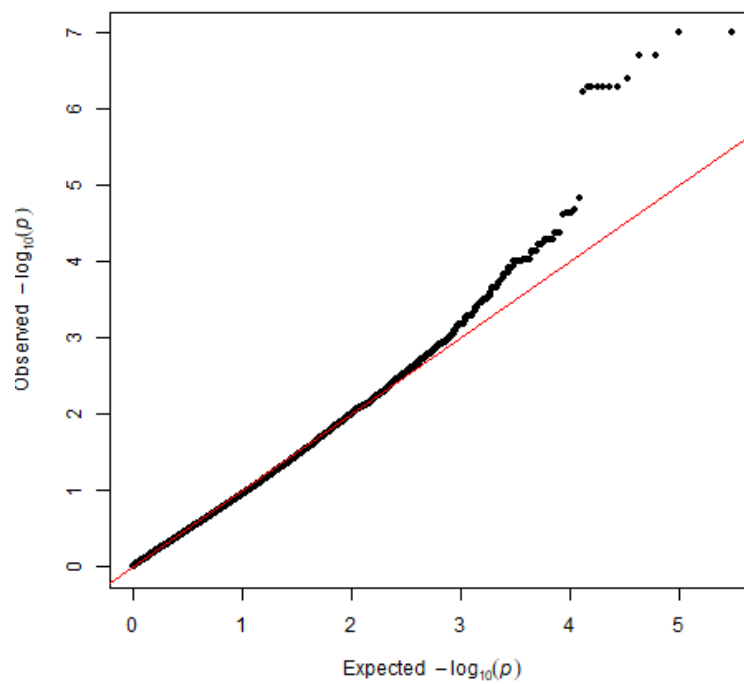


Figure 6.11 QQ plot for C3 at 8M.

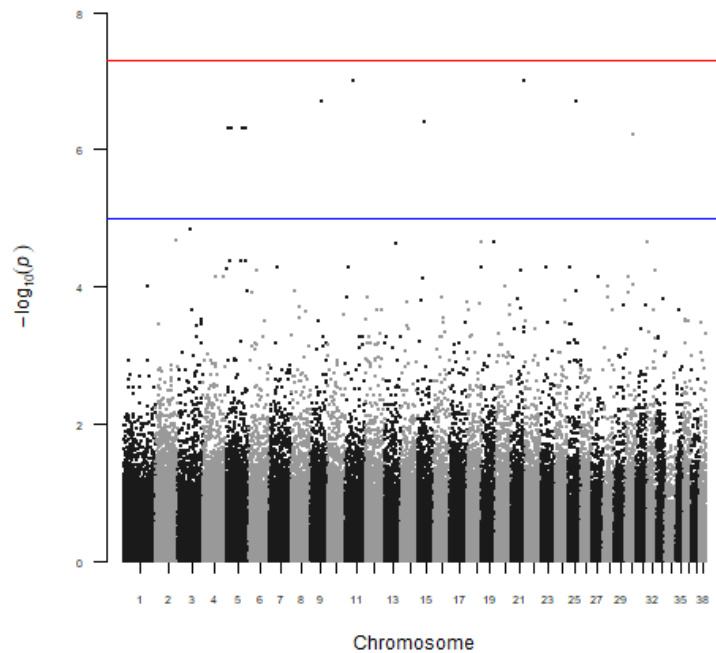


Figure 6.12 Manhattan plot for C3 at 8M.

After all genes identified in the regions of interest were submitted to DAVID, the highest scoring functional annotations/terms/diseases were:

- Alcoholism, substance-use disorders (enrichment score 2.09)
- Fatty acid metabolism (enrichment score 2.07)
- Sensory perception of smell, olfactory receptors (enrichment score 1.32)
- Depressive disorder, bipolar disorder, schizophrenia (enrichment score 1.19)

A summary of genes in the genome regions of interest can be found below in Table 6.10. Associations to human disorders include depression, Alzheimer's disease, Schizophrenia, Autism and PTSD (Post Traumatic Stress Disorder). There was also a link to canine herding behaviour reported.

Table 6.10 Associated behavioural phenotypes for C3 at 8M. The chromosome number, base pair position (Mb) of SNP(s) and the nearest gene are reported for the 20 SNPS with lowest P-values.

Chromosome	Mb of SNPs with lowest P-values	Nearest gene(s)	Associated behavioural phenotype
2	79.49	RNF186, TMCO4	Alzheimer's disease and dementia (Mez <i>et al.</i> , 2017; Sherva <i>et al.</i> , 2023)
3	46.22	MCTP2	Major depression (Verma <i>et al.</i> , 2008), ADHD (Kweon <i>et al.</i> , 2018), Alzheimer's disease (Harper <i>et al.</i> , 2022)
5	8.45 – 20.01	95 genes in region including olfactory receptors OR6, OR8, OR10 gene families	See below
5	53.95 – 75.06	165 genes in region	See below
9	39.88	ASIC2	Pain behaviour (Staniland and McMahon, 2009), response to aversive stimuli (Price <i>et al.</i> , 2014), reaction times (Davies <i>et al.</i> , 2018), major depressive disorder (Howard <i>et al.</i> , 2018; Levey <i>et al.</i> , 2021; Als <i>et al.</i> , 2023; Meng <i>et al.</i> , 2024), canine herding behaviour (Shan <i>et al.</i> , 2021)
11	28.64	KDM4C	Bipolar disorder (Wang <i>et al.</i> , 2010) Schizophrenia (Li <i>et al.</i> , 2017; Greenwood <i>et al.</i> , 2019; Ikeda <i>et al.</i> , 2019) autism spectrum disorder (Kato <i>et al.</i> , 2020)
13	46.01	USP46	PTSD (Kim <i>et al.</i> , 2007), behavioural despair/ immobility test in mice (Tomida <i>et al.</i> , 2009), bipolar disorder and schizophrenia (Kushima <i>et al.</i> , 2010), major depression (Boo <i>et al.</i> , 2019), dementia (Fongang <i>et al.</i> , 2024)
15	27.42	LRRIQ1	-
18	55.29	SCGB1A1	Suicide attempts and bipolar disorder (Sandberg <i>et al.</i> , 2022), memory performance (Eissman <i>et al.</i> , 2024)
19	45.01	LRP1B	Cognitive restraint (Cornelis <i>et al.</i> , 2014), Alzheimer's disease (Shang <i>et al.</i> , 2015; Kulminski <i>et al.</i> , 2022), eating behaviour (Ignatieva <i>et al.</i> , 2016), infant cognitive

			ability (Sun <i>et al.</i> , 2020), Schizophrenia, bipolar disorder or major depressive disorder (Blokland <i>et al.</i> , 2022), dementia (Real <i>et al.</i> , 2023)
21	47.63	No known gene	-
25	29.62	ELP3	Suicide ideation (Perroud <i>et al.</i> , 2012), long-term olfactory memory (Yu <i>et al.</i> , 2018)
30	30.69	No known gene	-
32	2.23	CNOT6L	-

Running 260 genes from the two regions on Chromosome 5 through DAVID, to search for KEGG and REACTOME pathways and UP Tissue gave the following results. Pathways with the highest enrichment score (3.24) were sensory perception (18 genes), olfactory transduction (15 genes), olfactory signalling pathway (12 genes), expression and translocation of olfactory receptors (11 genes). The Chromosome 5 regions contain several genes from the olfactory receptor families: Olfactory receptor family 10 – 2 genes, Olfactory receptor family 8 – 3 genes, Olfactory receptor family 6 – 6 genes. One hundred and six genes in the Chromosome 5 regions are expressed in the brain.

6.4.6.2 Component 5 - Anticipatory Excitement

Component 5 (C5) had a genomic-based heritability estimate of 0.35 (± 0.49) and a pedigree heritability estimate of 0.09 (± 0.04) at 8M. The top 20 significant SNPs associated with C5 were found for a range of chromosomes (Table 6.11). SNPs were located on Chromosomes 1, 3, 7, 8, 9, 12, 14, 21, 22, 24, 26, 29 and 31. A cluster of three SNPs were found on Chromosome 31 between 0.3 and 9.6Mb and two SNPs in the region of 34.3-35.4Mb.

Table 6.11 Top 20 significant SNPs for C5 at 8M. Chromosome number, SNP, base pair position and P-value are reported.

Chromosome	SNP	Region (bp position)	P-value
1	TIGRP2P17199_rs8966927	22189988	0.00009
3	BICF2S23210420	46220253	0.00010
7	BICF2P666484	38406698	0.00004
7	TIGRP2P103126_rs9156626	72038443	0.00008
8	BICF2S2298543	54099763	0.00005
9	chr9_45011893	45011894	0.00003
12	TIGRP2P159498_rs8802930	18391564	0.00013
14	BICF2S23653011	44008936	0.00014
21	chr21_6444541	6444541	0.00006
21	BICF2P1315874	13366876	0.00008
22	BICF2P635222	9872168	0.00004
24	chr24_5377386	5377386	0.00004
26	BICF2P655021	9782963	0.00003

26	BICF2P672331	23122504	0.00003
29	BICF2P341807	11558277	0.00008
31	BICF2P483487	291186	0.00004
31	BICF2P492048	3332126	0.00004
31	BICF2P489064	9637705	0.00004
31	BICF2P504025	34371151	0.00008
31	BICF2P485087	35409688	0.00004

The QQ plot (Figure 6.14) showed a deviation below the normal distribution line and no SNP effects exceeded the line denoting suggestive significance on the Manhattan plot (Figure 6.15). Results suggest that there were no SNPs with significant effect on C5 at 8M. Therefore, an in-depth exploratory analysis of the 20 significant SNPs was not undertaken.

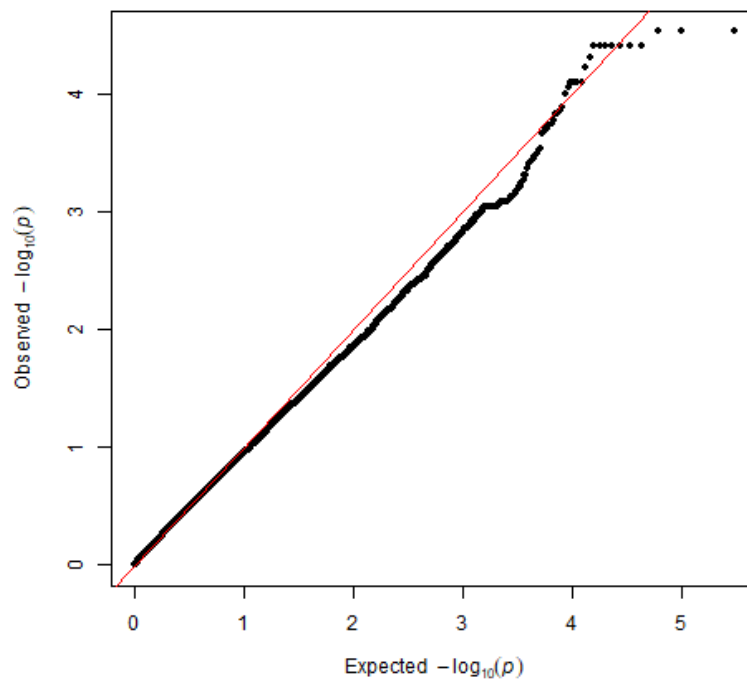


Figure 6.13 QQ plot for C5 at 8M.

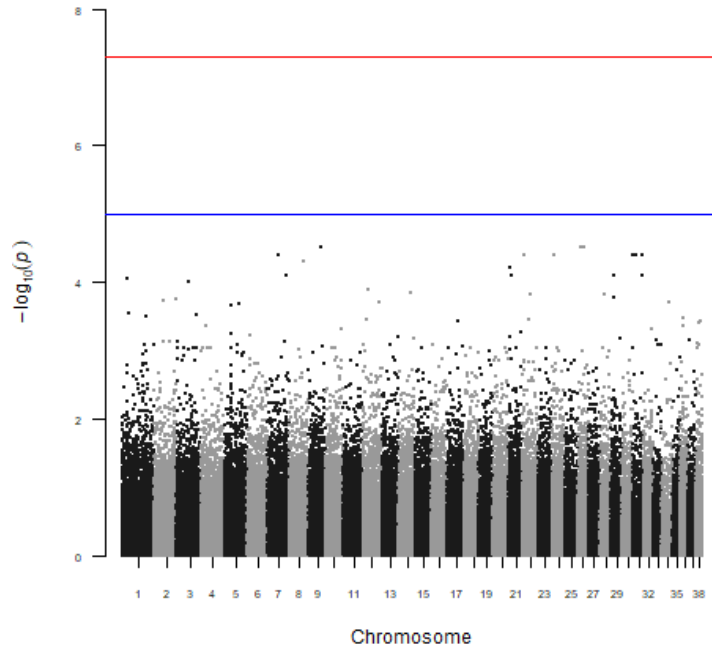


Figure 6.14 Manhattan plot for C5 at 8M.

6.4.6.3 Component 8 – Focus

Component 8 (C8) had a genomic-based heritability estimate of 0.91 (± 0.46) and a pedigree heritability estimate of 0.03 (± 0.03) at 8M. The top 20 significant SNPs associated with C8 were found for a range of chromosomes (Table 6.12). SNPs were located on Chromosomes 6, 7, 9, 10, 11, 12, 13, 14, 16, 20, 22, 23, 28 and 31. Two SNPs were found on Chromosome 31 between the region of 14.9-16.1Mb and another at around 35.9Mb

Table 6.12 Top 20 significant SNPs for C8 at 8M. Chromosome number, SNP, base pair position and P-value are reported.

Chromosome	SNP	Region (bp position)	P-value
6	BICF2P558681	35249206	0.00033
7	BICF2G63084664	67248001	0.00013
9	chr9_48412191	48412191	0.00013
10	BICF2P748308	5339311	0.00011
10	BICF2S23252418	42224528	0.00026
11	BICF2G630294248	15548838	0.00026
12	BICF2P450648	29025861	0.00028
12	TIGRP2P170199_rs8764009	60135589	0.00011
13	BICF2G630612317	19350027	0.00013

14	BICF2P992084	9824859	0.00031
16	TIGRP2P216815_rs9147835	42693172	0.00021
20	BICF2S23721001	7263804	0.00011
20	chr20_57625180	57625180	0.00013
22	BICF2G630324904	24102641	0.00033
23	BICF2G630375706	29301880	0.00013
23	BICF2P1342121	3953006	0.00021
28	BICF2G630265152	34318352	0.00022
31	BICF2S23553129	14890800	0.00026
31	BICF2G630733394	16068164	0.00016
31	BICF2S23612746	35932222	0.00026

The QQ plot (Figure 6.16) demonstrated a deviation below the normal distribution line, and no SNPs were noted to be above the line denoting suggestive significance in the Manhattan plot (Figure 6.17). Results suggest that there were no SNPs with significant effect on C8 at 5M. An in-depth exploratory analysis of the 20 significant SNPs was not undertaken.

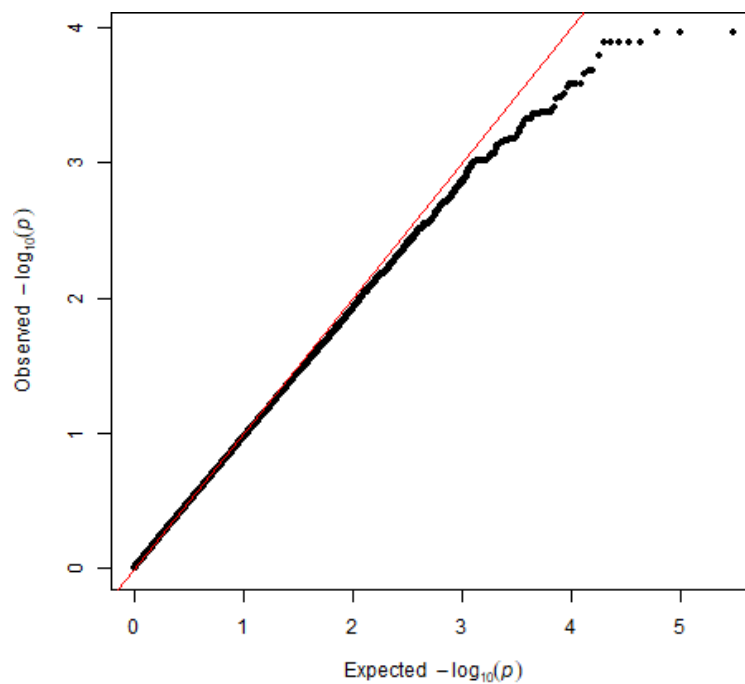


Figure 6.15 QQ plot for C8 at 8M.

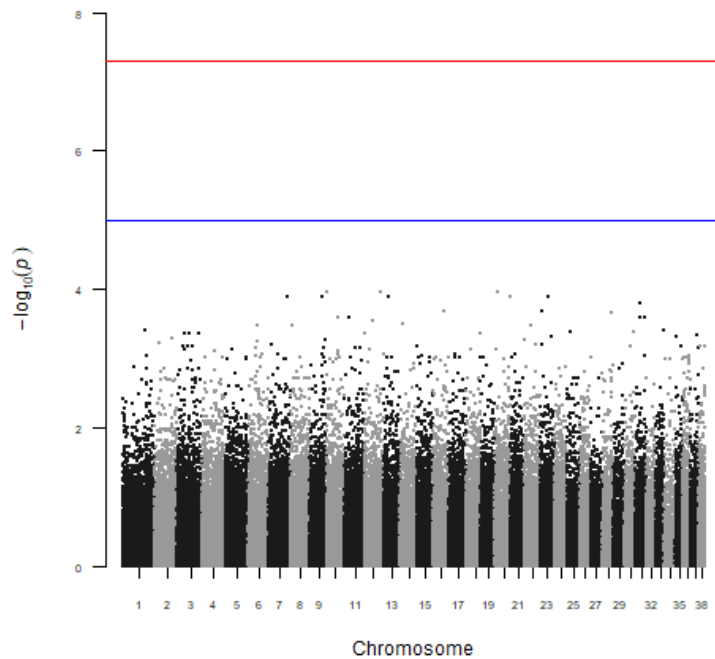


Figure 6.16 Manhattan plot for C8 at 8M.

6.4.7 GWAS results summary

There were similarities between the significant regions found between Separation (C7) at 12M and Anticipatory Excitement (C3) at 8M for Chromosomes 13 and 19. A region on Chromosome 13 was found at 47.72 Mb for C7 and 46.01 Mb for C3. For Chromosome 19, the region associated with C7 was found at 42.02 Mb and 45.01 for C3. There were also similarities in the human behavioural diseases/disorders that were associated with the genes of interest for both Separation at 12M and Anticipatory Excitement at 8M. Both components had significant genes associated with Alzheimer's disease, Autism, depression, bipolar disorder and abnormal eating behaviours (including anorexia nervosa and bulimia).

GWAS results for the 12M r-PWQ indicated three components to have a SNP based heritability estimate above 0.1. When these results were explored further, QQ and Manhattan plots suggested only Anticipatory Excitement had associated SNPs reaching a suggestive level of

significance. GWAS results for 8M r-PWQ indicated three components to have a genomic-based heritability estimate above 0.1. When these results were explored further, QQ and Manhattan plots suggested only Separation had associated SNPs reaching a suggestive level of significance.

6.5 Discussion

This study compared the pedigree-based and genomic-based heritability estimates of the r-PWQ2 for data collected at 8M and 12M using genotype data for 96 dogs. Results found the highest genomic-based heritabilities to be for Excitement at both 8M and 12M, Energy behaviours at 12M, and Separation and Focus behaviours at 8M using the r-PWQ2. These genomic-based heritabilities were higher than the pedigree-based estimates reported in Chapter 4. Following a review of QQ and Manhattan plots, SNPs approaching significance were shown for C7 (Anticipatory Excitement) at 12M and C3 (Separation) at 8M. The dog is a good model species to aid our understanding of human genetics due to the incidence of behaviour issues and diseases seen in breeds (Wayne and Ostrander, 1999). Amfim *et al.* (2024) reported there to be 607 dog genetic phenotypes that could be used as a predictive model for human disease. Consequently, the present study has explored links to extracted genes of interest and behavioural phenotypes in humans.

The sample size for this study was limited by funding available, and therefore the study was undertaken as an exploratory analysis. The sample size could explain the non-significant heritabilities found through genomic-based estimates for several components. However other studies have reported using similar numbers of dogs for genotyping. For example, Bellamy *et al.* (2018) studied 104 Havanese dogs and Sarviaho *et al.* (2020) reported data for 124 dogs. A study looking at

OCD-like tendencies in dogs used data from 40 dogs (Cao *et al.*, 2021), suggesting that small samples could be used for genotype studies. GDUK are sequencing around 3000 dogs are part of a genetic project called 'Born to Guide' (Guide Dogs, 2020b), and as such there will be a considerably larger pool of genotype data for comparisons in the future. It will take many years before these data are available, however initial findings from the present study suggest that it may be possible to calculate genomic-based heritability estimates for some components utilising the r-PWQ2.

Friedrich *et al.* (2019) undertook a similar study to this one and used questionnaire data (C-BARQ) to compare pedigree and genomic-based heritability estimates. Dogs included in the analysis varied across behaviour traits, with a range of 906 to 1038 dogs for pedigree-based heritability estimates and between 638 and 729 dogs for genomic-based estimations. Friedrich reported more similarities between pedigree-based and genomic-based heritabilities than found in the present study (e.g. pedigree-based heritability of 0.23 and genomic-based heritability of 0.17 for 'human-directed playfulness'). A reason for this may be the larger sample size used by Friedrich. Some traits reported by Friedrich showed differences between pedigree and genome methods (e.g. stranger-directed interest pedigree-based heritability of 0.10 and genomic-based heritability 0.01), like differences reported in the present study.

It is important to consider that owner-reported questionnaires are prone to high levels of bias (Bellamy *et al.*, 2018). The phenotype has the potential to be more variable, and thus impact any GWAS findings. Xue *et al.* (2021) noted challenges of using a GWAS for data that changes over time (e.g. behavioural displays). The r-PWQ2 questionnaire had different component groupings between the three time points (5M, 8M and 12M), so the questionnaire data is not directly comparable. This may explain some of the differences seen between the genomic-based

heritabilities found between 8M and 12M data. It would be interesting to explore differences with the 5M data in the future and for other data methods such as observational data. Due to the high pedigree-based heritabilities found for observational data collected in Chapter 4, future studies could explore differences in the pedigree-based and genome based heritabilities using this data.

When studying behavioural genetics, it is essential to consider the environment and how the experiences and upbringing of a dog effects behaviour. Previous work has shown the impact of rearing and environmental influences on the development of behaviour in a group of guide dog puppies (Harvey *et al.*, 2016) and findings from both Chapters 2 and 3 have indicated multiple factors influence puppy behaviour in the present population. Due to the sample size, the present study was unable to account for different environmental factors in the genome-based heritability models. QQ plots also indicated that the model may have over-fitted for breed effects. In future studies a separate model for each component should be explored due to the impact that breed may have on dog behaviours (MacLean *et al.*, 2019).

Following QC controls, approximately 160,000 SNPs remained from the 230,000 sequenced. Whilst it may appear that many SNPs have been lost, other studies have reported considerably lower numbers of SNPs for GWAS studies. Matsumoto *et al.* (2023) used 230K SNPs (the same as used in the present study) and following QC measures worked with 124,675 SNPs in German Shepherd Dogs and 154,340 SNPs in Labrador Retrievers, similar to the final number of SNPs used in the present study. Wang *et al.* (2021) reviewed genetic associations for canine hip dysplasia in German Shepherd Dogs using only 62,089 SNPs from the 180k Illumina chip. That study reported multiple loci associated with the disease dependent on the population. Similarly Friedrich *et al.* (2019) undertook a GWAS on 78,088 SNPs. The number of SNPs included in final analyses will depend on the

thresholds set, such exclusion of SNPs that are missing in a high proportion of individuals (--geno), exclusion of individuals with a high rate of genotype missingness (--mind), removal of SNPs with low minor allele frequencies (--maf) and Hardy Weinberg (--hwe). The present study used a --maf of 0.01 which has been reported as a common threshold by other studies (Jenkins *et al.*, 2021). Jenkins *et al.* (2021) similarly used a genotype call rate of 95%, a SNP call rate of 97% and the Hardy–Weinberg equilibrium $P < 5 \times 10^{-5}$. Therefore, the number of SNPs analysed in the present appears to be acceptable and comparable to other studies.

GWAS studies have successfully found associations between individual behaviours and particular genome regions in several different dog breeds (Joshi, 2005; Persson *et al.*, 2016; Ilska *et al.*, 2017). Bhowmik *et al.* (2024) found significant regions on multiple chromosomes for Social fear, Non-social fear and Startle in a population of 615 dogs over 18 months of age in the USA from varying breeds. Handegård *et al.* (2023) noted regions on Chromosome 17 to be linked to firework fear and noise reactivity in Standard Poodles, with the present study reporting SNPs on Chromosome 17 to be associated with excitement behaviours at 12M. Sarviaho *et al.* (2020) reported associations between loci on Chromosome 11 in a cohort of 124 Finnish Great Dane dogs, with associations made to genes that are reported to be involved in anxiety regulation, with the present study reporting SNPs on Chromosome 11 to be associated with Separation and Focus behaviours at 8M.

There were similarities found between the recent study by Bhowmik *et al.* (2024) and the SNPs extracted in the present study. Social Fear SNPs were associated with SNPs found in the present study for Excitement at 12M on Chromosome 13 (42Mb compared to 47Mb) and for Chromosome 5 (15 and 59Mb compared to 20 and 53Mb), Chromosome 13 (42Mb compared to 46Mb) and Chromosome 30

(26Mb compared to 30Mb) for Separation at 8M. For the startle response, similarities to SNPs associated with C7 (Anticipatory Excitement) at 12M on Chromosome 7 (62-66Mb compared to 72Mb) and Chromosome 22 (32Mb compared to 30Mb) were present. For the startle response, SNPs on Chromosome 2 at 79Mb were reported by Bhowmik *et al.* and in the present study for C3 (Separation) at 8M. Similarly, regions on Chromosome 9 (28Mb compared to 39Mb), Chromosome 21 (50Mb compared to 47Mb) and Chromosome 32 (23Mb compared to 22Mb) were associated with C3 at 8M at the startle response by Bhowmik *et al.* It is important to note that there are differences in the methods used between the present study (using a questionnaire) and the study by Bhowmik *et al.* (using a practical assessment). The cohorts of dogs were different; for dogs of different ages, breeds and purpose, yet there are still some comparisons that can be made between the findings from both studies and further evidence to suggest fear and excitement related behaviours in dogs may have a genetic influence.

6.5.1 Excitement associated behaviours

The highest genomic-based heritabilities were for excitement-related behaviours at 12M. General excitement (C6) had a genomic-based heritability of 0.43, and Anticipatory Excitement (C7) a heritability of 0.44 at 12M, whilst at 8M Anticipatory Excitement (C5) had a heritability of 0.35. The pedigree-based heritability estimates from Chapter 4 were below 0.1 for these components and previous work during the development of the PWQ showed Excitement to be impacted by sire (11.2% of variation) and litter (10.1% of variation; Harvey, 2014). Excitement and hyperactivity-related behaviours are undesirable for guide dogs (Craigon *et al.*, 2017) and these behaviours have been associated with guide dog success at 8M (Hunt *et al.*, 2020).

Following the GWAS, significant genes associated with the Anticipatory Excitement component at 12M were correlated to human behavioural disorders such as ADHD and bipolar disorder. Potential candidate genes associated with alcoholism and tobacco-use disorder in humans were also found. There is evidence for positive genetic correlations between substance-use disorder and psychiatric disorders in humans (Gelernter and Polimanti, 2021). Studies of human personality have found neuroticism (one of the big five personality traits) to be linked to depression and anxiety (Gupta *et al.*, 2024). Links have also been made to neuroticism and substance-use disorders (Kotov *et al.*, 2010), suggesting that personality influences an individual's likelihood to display more extremes of behaviour. It is thought that the heritability of neuroticism (estimated to be around 35-75%) could explain a genetic basis for mental health and psychiatric disorders in humans (Sanchez-Roige *et al.*, 2018). Sarviaho *et al.* (2019) found associations between traits on Chromosomes 7 and 20 to be associated with fearful behaviours in dogs, with overlap to human neuropsychiatric conditions. Similarly, the present study found genes on Chromosomes 7 and 20 to be associated with Anticipatory Excitement behaviours. These behaviours may arguably be on the opposite end of the spectrum to "fearful behaviours", but associations between the two extremes of behaviour can be made. Relating findings to guide dog puppies in the present study and comparing to the literature in humans, it is suggested that dogs exhibiting greater anticipatory excitement may also be more prone to display anxiety, hyperactivity and obsessive like behaviours.

The GWAS results suggested some associations to feeding behaviours and eating disorders in humans. Research in Labradors has demonstrated the significance of the POMC gene for obesity in dogs (Raffan *et al.*, 2016), with a higher frequency of the POMC deletion in assistance dogs, including guide dogs. Labradors are a common breed used by assistance dog organisations, with the breed known to show high food motivations (Wallis *et al.*, 2024), a trait useful for training

dogs. Findings from 12M GWAS showed suggestive significance for the RTF2, GCNT7, U6 genes on Chromosome 24, which are very close in location to the MC3R gene. The MC3R gene has been associated with anxiety and anorexia nervosa (Cho *et al.*, 2023) in mice. Interestingly there are also reported links to hyperactivity and anorexia nervosa in humans, suggesting that feeding behaviour and activity levels could be associated (Kohl *et al.*, 2004; Scheurink *et al.*, 2010). The present study suggests a similar link between canine excitement which could be explored further.

A review of canine genetics and epidemiology of dog behaviour undertaken by Amfim *et al.* (2024) discussed associations between ADHD in humans and canine impulsivity and then between obsessive-compulsive behaviours in both species. The present study found links between the excitement related behaviours displayed in dogs and ADHD in humans, specifically related to Chromosome 19. There are reports that ADHD-like behaviours can manifest in dogs (Hoppe *et al.*, 2017), and similarities are noted between the neurotransmitter pathways that are present in human ADHD and ADHD-like behaviour in dogs (González-Martínez *et al.*, 2024). Recently there has also been the creation of the Dog ADHD and Functionality rating scale (Csibra *et al.*, 2024), suggesting that ADHD-like behaviours are of increasing interest within dog behaviour research.

6.5.2 Separation associated behaviours

Separation at 8M approached significance during the GWAS with a genomic-based heritability of 0.62. Separation-related behaviours are a presentation of anxiety/fear related behaviours (Ogata, 2016) which have been shown to be predictive of guide dog success at 8M (Hunt *et al.*, 2020). Separation anxiety is one of the most common behavioural problems seen in dogs (Meneses *et al.*, 2021), with noted impacts on

welfare and potentially lifespan. As such this behavioural presentation is important to understand within the dog.

In the present study, for the top 20 significant SNPs associated with Separation, eight were found on Chromosome 5. Bel-Fenellós *et al.* (2023) found associations with cognitive behavioural profiles and a deletion present in Chromosome 5. When correlating genes associated with Separation in dogs and disorders in humans, similarities were found to depression, bipolar disorder, schizophrenia, autism spectrum disorder, and PTSD. More severe behaviour responses associated with separation anxiety in dogs can include self-injury (Meneses *et al.*, 2021). There were associations between the r-PWQ2 Separation behaviours and major depression and suicidal related behaviours in humans (Verma *et al.*, 2008; Perroud *et al.*, 2012; Boo *et al.*, 2019; Sandberg *et al.*, 2022). Bjerknæs (2019) suggested that maladaptive behaviours could be present in dogs, where fearful dogs also are prone to showing higher levels of other behaviours such as separation and aggression. Maladaptive behaviours in humans can include self-injury, aggression, substance use and disordered eating (Lanfredi *et al.*, 2021). These behaviours show strong links to the behavioural phenotypes associated with genes linked to Separation found in the present study. Results suggest that the presentation of extreme separation behaviours could be associated in some way with human displays of anxiety and mental health issues, which may also manifest within dog behaviour, although this has yet to be explored further.

Olfaction and links to behaviour is a topic to consider. Dogs are known for their acute sense of smell, which is highly specialised and sensitive when compared to humans (Horowitz *et al.*, 2013; Kokocińska-Kusiak *et al.*, 2021). This ability may be used in many different working roles such as scent and medical detection dogs, including for example the ability to detect explosives (Gazit and Terkel, 2003) and patients with lung cancer (Amundsen *et al.*, 2014). Research in humans suggests

that olfaction is especially critical in childhood and is thought to play a role in the development of social relationships (Roberts *et al.*, 2020). For Separation, there were significant associations with genes on Chromosome 25 and links to long-term olfactory memory in humans (Yu *et al.*, 2018). Olfactory and sensory processing can contribute to behavioural differences between individuals (Hoover, 2010; Aron *et al.*, 2012) and olfaction has been associated with psychiatric diseases in humans (Huckins *et al.*, 2013). This poses questions as to how individual senses may impact behavioural differences observed in dogs and is an area requiring further study. This may be possible using data collected during the Odour Discrimination task in Chapter 4 alongside genotype data.

Studies in dogs have found associations between fearfulness and genomic regions. Sarviaho *et al.* (2020) noted fearful behaviours to be present on Chromosome 11. The present study identified genes on Chromosome 11 associated with Separation, which are associated with the fear and anxiety spectrum (Schiele *et al.*, 2020). Significant SNPs were also found on Chromosome 15 and 18 for r-PWQ2 Separation. Research undertaken by Zapata *et al.* (2016) reported IGF1 loci variants (found on Chromosome 15) and genes on Chromosome 18 to be associated with Separation-related behaviours in dogs using the C-BARQ questionnaire. The Separation questions used in the r-PWQ show strong correlations to the C-BARQ questions (Hunt *et al.*, 2020). Findings support the genetic influence on anxiety related behaviours in dogs.

6.5.3 Future research

The advancements of GWAS are evolving to include imputation which is reported to be a “cost-effective method to increase the density of extant lower-density array datasets” (Jenkins *et al.*, 2021). Imputation predicts missing genotypes in a data set of genotyped individuals using

individuals with higher density sequencing such as whole genome sequences (Jenkins *et al.*, 2021). Hayward *et al.* (2019) states that through imputation, it was possible to increase the number of variants reviewed in analysis from 173,000 to 24 million. Whilst not included within this chapter, there is scope for further work to include imputation of the genotypes used from these 96 dogs due to the whole-genome sequencing undertaken by GDUK as part of Born to Guide (Guide Dogs, 2020b). In addition, this study only focussed on the top 20 significant SNPs for each component due to the small sample size and pilot nature of the analysis. Therefore, future studies with a larger sample size and more in-depth genotype analysis could uncover more significant SNPs associated with behavioural phenotypes.

GDUK are aiming to collect and sequence data for around 3,000 dogs, enabling a multitude of studies to be undertaken on behaviour. It is suggested that a future study incorporate all three time points for the r-PWQ2 and include outcome (qualification as a guide dog or withdrawal). A clear distinction between the pure and crossbreed dogs could be seen during the quality controls for this GWAS study. A greater sample size would also enable a deeper exploration into breed effects and differences between purebred dogs and crossbred dogs. In addition, other environmental factors could be included in the model and each component could have an individualised model based on factors significant to the behavioural phenotype. Findings from Chapter 3 highlighted that the 5M r-PWQ2 time point had the highest pedigree-based heritabilities when compared to the 8M and 12M time points. Due to the timing of available funding, it was not possible to select dogs with the 5M r-PWQ2 for genotype sampling. Future studies could review a sample of dogs with data and all three time points and compare pedigree-based and genomic-based heritabilities.

This study has reviewed the human literature to find associations to behavioural phenotypes and genes of interest. As more research is

undertaken on humans and other species, future work could update analysis of specific genes to determine whether any further links can be made between species or if further research has been undertaken in dogs. Despite the small sample size in the present study, following a GWAS, two r-PWQ2 components demonstrated approaching significance in a population of guide dog puppies. Arguably the behavioural repertoire of guide dogs is smaller than that seen within other breeds or pet dogs, which will be due to extensive selective breeding for behaviour over many generations. Therefore, the findings from this pilot study provide a promising foundation for future studies comparing canine and human behaviours.

6.6 Conclusion

Dog behaviours have evolved alongside humans through the domestication of the species, which is why the present study, and many others, have compared findings in the dog to those in humans. The present study explored differences between pedigree-based and genomic-based heritability estimates, answering the research question. A GWAS identified associations between dog behavioural displays and obsessive and depressive disorders in humans such as ADHD, alcoholism, depression, eating behaviours and Schizophrenia. Separation behaviours are a manifestation of anxiety behaviours, which could show associations to depressive like behaviours. There were links to genes associated with separation behaviours in dogs and depression, Schizophrenia and bipolar disorders in humans. Anticipatory excitement behaviours could show associations to obsessive and substance-use related disorders. Excitement behaviours displayed in dogs showed correlations to alcoholism and ADHD in humans. The above findings provide several potential new opportunities to explore and aid our understanding of canine behaviour. This study has laid the foundation for future GWAS studies at GDUK into behavioural traits, and ultimately links to qualification as a guide dog.

Chapter 7. General discussion and conclusions

The aims of this thesis were to determine the genetic influences on guide dog behaviour through pedigree and genomic-based heritability estimates for data captured through questionnaires and a behavioural assessment. Appropriate behavioural responses are a necessary component contributing to a guide dog successfully completing training. Therefore, it is important to understand to what level behaviour is influenced by genetics and thus could be improved by selective breeding. The present study first set out to determine if a refined version of the puppy walker questionnaire (Hunt *et al.*, 2020) remained valid for use in a population of guide dog puppies, where the raising process had changed from when the questionnaire was developed. Following further development of the r-PWQ to the r-PWQ2 (Chapter 2), the study aimed to determine the heritability of the behavioural components derived from the questionnaire (Chapter 3). Previous work at Guide Dogs UK (GDUK) found heritabilities of behaviours observed in young puppies during a behavioural assessment (Evans *et al.*, 2015). The present study then undertook a validated behavioural observation assessment developed for assistance dog puppies (Bray *et al.*, 2020) and determined the heritability of the assessment (Chapter 4). The reliability of the assessment was reviewed (Chapter 5) and a pilot study utilising machine learning (ML) undertaken to determine whether artificial intelligence (AI) capabilities could aid in accurately reporting puppy behaviour. Finally, genomic-based heritabilities were determined using data from the r-PWQ2 for 96 dogs (Chapter 6), to compare to pedigree-based heritability estimates and to identify associated genomic regions using a genome-wide association study (GWAS).

Unacceptable behaviour accounts for most withdrawals from guide dog programmes internationally (Goddard and Beilharz, 1984; Serpell and Hsu, 2001; Arata *et al.*, 2010; Foraita *et al.*, 2024). The qualification rates were 51% for the latest cohort analysis undertaken at GDUK for

dogs born in 2020, with previous cohorts demonstrating qualification rates generally between 50 and 60%, but as low as 38% in 2019 (Moxon and Fellows, 2024). Blythe (2020) analysed the main reasons for withdrawal of dogs from the GDUK training programme between 2001 and 2009 and found 25% of dogs withdrawn for behavioural reasons were for 'distraction' and 18% for 'suspicion'. Similarly for dogs withdrawn between 1995 and 2012, 68% of dogs were withdrawn for behavioural reasons, with 22% for 'high distraction' and 16% for 'suspicion' (Evans, 2014). These are similar to findings reported by other guide dog schools internationally, including Japan (Arata *et al.*, 2010), the USA (Serpell and Hsu, 2001; Ennik *et al.*, 2006) and Australia (Goddard and Beilharz, 1982).

Since almost 50% of dogs in guide dog training programmes do qualify as guide dogs, even a slight increase in success rates would provide more dogs to support people with vision impairment and potentially require less dogs to be bred by organisations. This in turn could result in financial/resource savings as there would be fewer dogs to raise and support throughout their lives. The average lifetime cost of a guide dog in 2008 was reported to be around £35,000 (Wirth and Rein, 2008), with figures now expected to be approaching nearly double the costs due to inflation. Increasing behavioural success could also improve welfare since dogs unsuitable for guiding work are commonly reported to display fearfulness and anxiety-related behaviours. Dogs that fail as guide dogs require rehoming as pet dogs, which is time-consuming and costly, therefore there are ethical benefits from breeding the least number of dogs required.

7.1 Previous studies undertaken at GDUK

Many studies have been undertaken in collaboration with GDUK over the past few decades in order to help improve understanding on behaviour and provide evidence to aid decision making on dogs, such

as whether they should progress into training or be withdrawn from the programme. Asher *et al.* (2013) published a behavioural assessment (Puppy Profiling Assessments; PPA) for 6-8 week old puppies during 2006 and 2007, reporting that five of the 11 stimuli were associated with success in guide dog training. Evans *et al.* (2015) then found aspects of PPA to be heritable (h^2 ranged from 0.02 for noise response to 0.25 for following the assessor). GDUK then implemented PPA as a decision-making tool within the organisation until 2019. However, over time more staff were involved with the assessments and the inter and intra-reliability of scoring was not regularly checked. Consequently, large fluctuations were observed in the data based on different assessor combinations (data not published). At a similar time, GDUK updated all training practices and implemented positive reinforcement training (PRT) throughout the organisation, so a re-validation was required to ensure that PPA was still fit for purpose. GDUK management decided that PPA should be stopped based on a cost vs benefit basis. PPA was trialled at another guide dog school more recently in the USA (a school also using PRT) and the results showed no links to guide dog success (Feng *et al.*, 2024). This may indicate that PPA was appropriate for the training methods at the time when developed, but that the assessment would need updating to be relevant for current GDUK operational practices.

In parallel to the work to develop PPA, Harvey (2014) reviewed data collected during puppy raising and developed the Puppy Walker Questionnaire (PWQ). The PWQ was published (Harvey *et al.*, 2017) and later was shown to be predictive of guide dog success (Hunt *et al.*, 2020). Subsequently, GDUK implemented the questionnaire within the organisation, which has formulated some of the data used in this thesis. GDUK also actively uses the questionnaire to monitor puppies in the programme and to aid in reviewing potential breeding dogs.

Evans (2014) undertook the first large scale genetic-based assessment of dogs at GDUK based on behavioural reports in puppy raising and training and found there to be few heritable behaviours. However, this is not surprising as the data used for analyses was shown to be unreliable. Results from the Epidemiology of Guide Dogs project (Asher *et al.*, 2016) found the assessor completing the assessment to be one of the greatest influences on behaviour reports when assessors scored dogs on a 1-4 scale for a subjective list of terms such as 'confidence' and 'suspicion'. Blythe (2020) also noted that "subjective assessments are the default process used by GDUK to evaluate behaviour", highlighting an issue with the type of data collected in the organisation. GDUK updated their behaviour evaluation system in 2021 to remove the use of scoring based on subjective labels.

Craigon *et al.* (2017) highlighted the importance of working and non-working behaviours to guide dog owners (GDO). From a GDO perspective, attentiveness, obedience, confidence and consistency of the dog's behaviour were the most important aspects of behaviour for a guide dog. Interestingly there was a significant variation in opinion between GDO, highlighting that the organisation needs to produce a range of dogs for GDO. For example, some GDO enjoyed their dogs being attentive and focussed on them out of harness, others found this to be a negative characteristic. Similarly, some GDO enjoyed having a more excitable dog, whilst others found this unsettling. Linked to this work, Caron-Lormier *et al.* (2016) analysed working dogs removed from the guiding role between 1994 and 2013. Of the 7500 dogs removed, 17% were withdrawn for behavioural issues, with the main reasons including environmental anxiety and fear/aggression. This highlights the impact that anxieties and fear-related behaviours can have not only on the likelihood of a dog qualifying as a guide, but also the ability of the dog to continue supporting their GDO. Early removal and withdrawal of guide dogs is distressing for the GDO and requires additional resources from the charity to support the partnership.

Blythe (2020) and Harvey (2014) explored behavioural responses of dogs in training and during puppy raising using behavioural assessments at GDUK. Harvey (2014) developed the mobile puppy test which was undertaken on a group of 93 dogs at five and eight months of age. Blythe (2020) developed the Controlled Adult Dog Assessment (CADA) for 111 adolescents between the ages of 12- and 14-months entering guide dog training. Due to the structure of GDUK and the widespread geographical location of puppies and dogs, it is not possible for GDUK to undertake a reliable practical assessment on all puppies, largely because of standardisation limitations and the resources required to undertake and analyse the assessments. GDUK therefore must primarily rely on questionnaire data collected from multiple individuals. Ensuring the data is reliable and reflective of the dog's behaviour is important to aid in accurate decision-making.

7.2 Reliability of behaviour data

Whether scientists are measuring personality, behaviour or temperament, the data is only valuable and of use if accurate and valid. Researchers have discussed the importance of reliability in dog behavioural studies (Brady *et al.*, 2018; Harvey, 2021) and yet published papers consistently fail to report the reliability of their studies which questions whether this analysis has been undertaken. Studies assessing dog behaviour tend to use one of two methods: either behavioural observations (e.g. situational or in a test/assessment), or a review of an individual over a range of time and contexts (e.g. questionnaires). There are different issues regarding reliability and validity between the two methods. As reported in Chapter 5, variation in the inter and intra-rater reliability of the behavioural observation assessment used in the present study depended on the measure reviewed and the assessor. Utilising machine learning provided some promising foundation results and indicated that the technology could be

used to measure puppy behaviour. Wider use of this technology at GDUK and in other dog behavioural studies could be explored.

Particularly within assistance dog organisations, there are multiple individuals involved in the care and raising of dogs. Organisations often rely on volunteers to raise puppies, known as puppy raisers, and these individuals have varying time-availability, expertise and knowledge on dog behaviour. The ability to report and interpret behaviour amongst the large volunteer base will vary, and it will be virtually impossible to remove all subjectiveness and bias. Yet the views of these individuals are very important; they have spent the most time with dog through puppyhood and arguably 'know' the puppy best. Previous work has reported the value of puppy raisers assessments of the dog's behavioural ability (Batt *et al.*, 2009; Harvey *et al.*, 2016) and future guide dog success. It has been suggested that matching puppy raisers and puppies based on the personality of both could lead to higher qualification rates (Wood, 2020). Mai *et al.* (2021) showed that socialisation and training predicted most puppy behaviours in an Australian cohort of guide dog puppies. They recommended that organisations should invest in resources to improve puppy raiser support, due to their large influence on training success. It is therefore important for GDUK to continually review the support provided to puppy raisers and reassess the predictive ability and reliability of the r-PWQ2 questionnaire.

To remove some of the subjectiveness present in questionnaires, researchers often look to capture behavioural observations from assessments/tests. The findings from Chapter 5 showed varying reliability for the practical assessment in Chapter 4 based on the experience of the individual completing the coding. A pilot utilising ML was then undertaken to explore whether the application could be reliable for measuring puppy behaviour from a video compared to human assessors. There are promising results for the use of ML in

behavioural analysis, including tracking data in real time for individual cows in a herd (AI, 2025). Yet it is difficult to completely eradicate human interpretation from analysis. Humans are required to train ML models, indicating specific behaviours or interactions that the computer should search for. There will still be a level of human influence, but this is required to ensure the ML is working correctly. One could argue that once the model has been trained by an individual (or group of individuals), the results are likely to be more consistent as human coders can be affected by fatigue and errors. Often behavioural observation assessments use a pre-defined ethogram, so biases on which behaviours to look for are already determined before the project has begun. It would be interesting to explore results gathered using computer analysis with no human input; questioning whether the computer would be able to detect subtle behaviours or trends that we as humans have not identified. Applications for using ML in the future have many possible avenues, including collecting live data on animals in facilities such as rehoming shelters, zoos or farms, to identify welfare issues as early as possible, in addition to providing rapid data from a testing environment enabling researchers to assess more individuals.

Environmental factors (as opposed to genetic factors) can have a significant impact on behaviour and, as much as possible, studies need to account for these factors. Results from Chapter 2 found that different environmental factors influenced behavioural responses, with factors either being external to the individual (e.g. presence of children, staff member supervising the puppy (PDA), the experience of the volunteer raising the puppy (Puppy Raiser) etc.) or intrinsic to the individual (e.g. sex, age at neutering, litter). Intrinsic factors may be more related to the animal's biology, e.g. hormone concentration due to neutering, sex of the animal, but are still considered as an environmental effect rather than a genetic effect. Determining differences between intrinsic and external factors could be an avenue to explore in more detail in future studies and researchers should ensure data is collected on a multitude

of potential influencing factors. The present study only accounted for a defined list of factors, but multiple factors not accounted for could also have affected the behaviours displayed. For example, maternal behaviours have been shown to influence puppy behaviour (Bray *et al.*, 2017), and other factors such as human personality and the dog vs human relationship could be explored. The factor with the greatest impact overall was shown to be the litter the puppy was raised in, suggesting that litter environment and genetic effects could be influencing the behaviours seen in this dog population. GDUK could consider collecting more information on the litter or behaviours when in the nest. Evaluation of such data might facilitate the development of intervention strategies to improve training outcome success.

The internal reliability of the r-PWQ2 components in this study ranged from 0.34 to 0.86 for all components at the three time points (5M, 8M and 12M). It is generally accepted that a Cronbach's alpha value of 0.7 and over suggests a reliable component (Taber, 2018). The present study found two components at 5M, three components at 8M and three components at 12M that had a Cronbach's alpha below 0.7, but half of these were above 0.6. The study could have been improved by removing the unreliable components from the analyses. A decision was made to include all components due to not understanding the genetic influences on behaviour, and a preference to include as many phenotypes as possible for analysis due to the smaller sample size than anticipated due to Covid-19.

7.3 Pedigree-based and genome based heritabilities.

To improve behaviour within assistance dog breeding programmes, it is essential to understand the genetic influences on behaviour (Hradecká *et al.*, 2015). If behaviours are primarily influenced by environmental effects, breeding programmes will not be able to influence the behaviours in future generations (Jensen and Andersson, 2005), so

organisations can focus on the modification of behaviour through training. The present study made comparisons between pedigree-based and genomic-based heritabilities for data collected using a questionnaire (the r-PWQ2). The study also compared heritabilities for data collected from a behavioural observation assessment to questionnaire data.

There were differences between the two pedigree-based heritability studies undertaken in this thesis. Questionnaire-based methods found the highest heritability to be 0.29 for the r-PWQ2 Anxiety trait at 5M. Results from the behavioural observation assessment demonstrated higher heritabilities (22 behaviours demonstrating heritabilities above 0.4). The r-PWQ2 questionnaires were completed for puppies of five, eight and 12 months of age. These puppies had already experienced several transitions, from the nest to the GDNC, to a puppy raisers home, and in some cases more than one puppy raiser home (although this was not examined as part of the study). Puppies would have been exposed to different environments, and this was not controlled for within the study. Subjects for behavioural assessments were younger puppies aged between six and 10 weeks. Puppies were raised in volunteer homes (or in whelping facilities at the GDNC) and had experienced one transition to the GDNC where they were housed during assessments. There were also similarities between the cohorts, all dogs were bred and raised by GDUK (dogs born between 2016 and 2023). All dogs were from the same breeds (Labradors, Golden Retrievers, and crosses of these two breeds) and the analysis was undertaken using similar heritability models. Despite these differences, correlations between the two data types were found, and heritabilities were reported using both methods. It would be possible for GDUK to select for behaviours based on data collected through behavioural assessments or owner-reported questionnaires (a method of data collection that is much easier and requires less resources).

Dog behaviour pedigree-based heritability estimates are commonly below 0.4 (Ilska *et al.*, 2017), which was confirmed in this study for r-PWQ2 heritability results and for some measures of behavioural assessments (57 of 79 measures). Hradecká *et al.* (2015) reported heritabilities to be on average much lower, at between 0.09 and 0.15 for 48 studies reviewed. In the present study, all pedigree-based heritabilities were below 0.1 at 8M and 12M, with higher heritabilities reported at 5M. Results suggest that GDUK could explore calculating estimated breeding values (EBVs) for the four components with heritabilities above 0.1 at five months of age (Anxiety, General Excitement, Energy and Body Sensitivity) due to the suggested genetic influence on the phenotype variation. Yet it is important to note that the 5M sample was smaller than the 8M and 12M sample, so the relationship matrix between individuals in the sample would have been smaller and may have influenced the results.

Some r-PWQ2 components had higher genomic-based heritability estimates when compared to pedigree-based estimates, particularly Separation and Energy. Closer examination of these results revealed the differences were most apparent at 8M. The genomic-based heritability estimate for Separation was 0.62 for puppies at 8M (0.05 for pedigree-based estimate) but this decreased to 0.05 for the genomic-based heritability at 12M (0.06 for pedigree-based estimates). The Energy component also displayed a similar pattern, although the results were lower (genomic h^2 0.19 at 8M and h^2 0.00 at 12M). This poses questions as to why the results could be so different between the two ages, for the same cohort of dogs. One possibility could be change in the variation in phenotype and environmental influence on the observed phenotype at this age. Behavioural differences are noted in dogs at approximately eight months of age (Asher *et al.*, 2020), which are further impacted by the presence of puberty (Moxon *et al.*, 2023). Asher *et al.* (2020) concluded that scores for Separation-related behaviour were 36% higher at adolescence (8M) than pre-adolescence (5M) and

post-adolescence (12M). These findings suggest that the behavioural observations seen at 8M (which also demonstrate a higher genomic-based heritability) may be more reflective of the variance seen in Separation behaviours rather than at other time points. There was a greater consistency between the 8M and 12M genomic-based estimates for excitement-related behaviours. At 12M, the associated Excitement component had a genomic-based heritability of 0.43 and 0.44, and at 8M this was 0.35.

A genome-wide association study (GWAS) is used to review genetic variations present in the genome to identify associations between genotypes and phenotypes (Tam *et al.*, 2019). A GWAS was undertaken during the present study to determine whether there were any common genomic regions associated with dog behavioural displays in data collected using the r-PWQ2. The GWAS utilises specific genomic information whereas a pedigree-based heritability reviews correlations between related individuals. Therefore, a GWAS provides a more accurate depiction of the genomic influences when compared with a pedigree-based estimate. Results from the GWAS in the present study indicated regions in the genome to be associated with Component 7 at 12M (Anticipatory Excitement) and Component 3 at 8M (Separation), with further associations found to behavioural phenotypes in humans.

Separation (genomic-based heritability of 0.62 at 8M) and Excitement behaviours (genomic-based heritability of 0.44 at 12M) in this guide dog population demonstrated potential associations to psychological disorders in humans. Both components exhibited similarities in the presence of genomic regions of interest on Chromosomes 13 and 19, and both had associations with Alzheimer's disease, autism, depression, bipolar disorder and abnormal eating behaviours (including anorexia nervosa and bulimia) in humans. There were also associations found to genes that have been linked to feeding behaviours and

olfaction, which is a potential avenue for further exploratory analysis. In dogs, Sarviaho *et al.* (2020) found fearful behaviours to be associated with genes on Chromosome 11, with this present study also finding genes on Chromosome 11 associated with Separation, which are linked to anxiety (Schiele *et al.*, 2020). Genomic regions of interest extracted from this study have also recently been reported to be associated with fear related behaviours in dogs (Bhowmik *et al.*, 2024).

In the present study, behaviours such as attachment/attention seeking and trainability showed low heritabilities at all three r-PWQ2 time points for both genome and pedigree-based estimates. There is a possibility that little genetic variation was found for these behaviours in the population or that variation has been reduced due to selection. Environmental influence is suggested to have had a larger impact on these behaviours, or these behaviours may be more difficult for handlers to interpret accurately when completing the questionnaires. This is useful for working dog organisations to understand, as training modules could be reviewed and updated to improve behaviours in these areas (with the suggestion that there can be little influence made by selective breeding).

A behavioural observation study was used to explore pedigree-based heritabilities for six tasks undertaken on a cohort of GDUK puppies. Twenty-two of 79 behavioural measures from six tasks were shown to be heritable above 0.4, indicating that 40% or more variation in the phenotype could be attributed to genetic influences. Heritabilities are generally reported to be below 0.4 for dog behaviour (Ilska *et al.*, 2017), therefore results from the present suggest higher heritabilities than commonly reported for dog behaviour. A principal component analysis (PCA) of the heritable behaviours extracted six components that were then compared to the 5M r-PWQ2. The component comprising of responses to the 'Novel Object' task was correlated to four 5M r-PWQ2 components (Separation, Anxiety, General excitement and Body

sensitivity). Results suggest that interaction and response to a novel item, the startle/surprise response, problem solving, and human sociability all may have a genetic influence in this population based on the pedigree-based heritabilities found. Studies have previously reported heritabilities for human sociability (Van der Waaij *et al.*, 2008; Howell and Bennett, 2011; Bray *et al.*, 2021a) and the startle response (Alberghina *et al.*, 2020; Bhowmik *et al.*, 2024) in dogs. Behaviours reported to be associated with anxiety were consistently found to be heritable throughout the present study, for both r-PWQ2 questionnaire data and for behavioural observation assessment data. These findings further support the possibility of selection for these behaviours in this population.

7.4 Anxiety-related behaviours

Anxiety-related disorders are commonly reported in humans, with up to one-eighth of the world's population thought to be affected by the disorder (Kumar *et al.*, 2013). At the beginning of the century, it was estimated that anxiety disorders in the United States could cost as much as \$100 billion per year (Kessler and Greenberg, 2002). The disorder also has a great impact on wellbeing. Similarly, within working dogs, there is an expense associated with anxiety behaviours and an impact on dog welfare. A dog exhibiting anxiety-related behaviours will often require more training input and may require withdrawal from the programme. Yet as stated by Harro (2018), anxiety is 'not easy to describe and to connect with observable human behaviour'. Therefore, if anxiety is not easily determined in humans when using visual cues, the potential to accurately observe and report anxiety in animals may be questioned.

Personality development and anxiety disorders have been extensively studied in humans. Kagan (1997) found that highly reactive infants (defined as becoming motorically active and distressed to the presence

of three stimuli) had a greater risk for developing anxiety disorders as adolescents or adults. Anxiety-related behaviours are observed to be present within multiple species, with model animal species such as mice and rats commonly used to explore the disorder (Harro, 2018). However, is it appropriate to use the same term, 'anxiety', for humans and other animals? Belzung and Philippot (2007) discussed qualitative differences between human and animal anxiety. The researchers reported that human anxiety depends on autonoetic consciousness (an ability to examine events from the past that project into the future), which is only thought to be present in great apes. Yet more recently horses have been found to have the ability to make conscious choices in an inhibitory task (Evans *et al.*, 2024). This suggests that human anxiety and understanding of the term is unlikely to be directly applicable to other species, however exploratory analysis in other species is required to aid our understanding.

Dog behavioural studies refer to 'anxiety' and 'fear' interchangeably, providing some confusion as to the correct terminology to use. Kagan (1997) described reactions to unfamiliarity as being part of an animal's temperament, which is potentially a much less subjective way to describe anxiety-associated behaviours. Goldsmith and Lemery (2000) note that the temperamental concepts of 'inhibition, shyness and fearfulness' overlap with the anxiety domain, suggesting anxiety encompasses a range of behavioural components. Following this work, Svartberg (2002) discussed how Playfulness, Curiosity/Fearlessness, Chase and Sociability were all related to a higher dimension known as the 'Shyness-boldness' dimension, with the dimension then shown to be consistent for individuals in three repeated tests over at least two months (Svartberg *et al.*, 2005). King *et al.* (2003) noted there to be no consensus on how fear was recognised and measured in dogs. The authors collected data on behavioural observations and biological factors such as cortisol concentration and heart rate with an aim to measure fear in dogs. Following the study, the authors describe PCA

components to measure 'response to novelty' and 'response to startling stimuli', where the components together may measure aspects of fear (King *et al.*, 2003). It might therefore be proposed that the 'General Anxiety' component measured in the r-PWQ2 be referred to as a 'response to uncertainty'. Behaviours associated with anxiety (or response to uncertainty) have been reported as reasons for guide dog withdrawal (Goddard and Beilharz, 1982; Serpell and Hsu, 2001; Asher *et al.*, 2016; Moxon, 2018; Blythe, 2020) and early retirement (Caron-Lormier *et al.*, 2016). The component is also consistent within different populations; PWQ development (Harvey, 2014), the r-PWQ development (Hunt *et al.*, 2020) and during the present study. The component was shown to have the highest pedigree-based heritability estimate at 5M and was shown to have high correlations to the heritable behaviours in the behavioural observation assessment.

The pedigree-based heritabilities for the response to uncertainty (Anxiety) component in Chapter 3 were low for r-PWQ2 data collected at 8M and 12M (h^2 0.02 and 0.08 retrospectively). In addition, no significant genomic-based heritability estimates were found for the component at 8M or 12M (h^2 0.00 at both ages). Data collected at the first r-PWQ2 time point, when puppies would have the least environmental impacts of all the ages, found a pedigree-based heritability estimate of 0.29. Data at 5M was not available for dogs analysed during the genomic study in Chapter 6. The behavioural observation assessment data collected in Chapter 4 indicated that response to uncertainty behaviours also demonstrated high heritabilities in puppies around seven weeks of age. The findings suggest that earlier measures of response to uncertainty are essential to capture for guide dog puppies. Interestingly, work undertaken prior to the present study did not show the same component to be predictive of guide dog success for data collated at 8M using the r-PWQ (Hunt *et al.*, 2020). The 5M data were also not available for this cohort as on implementation of the PWQ at GDUK, Dr Harvey suggested the 8M and

12M questionnaires would be more useful to the organisation from a practical decision-making point of view (N. Harvey 2020, *personal communication*, 29 July 2020). The eight and 12M data points are impacted by a period of juvenile development and puberty, and due to these influences, may not be reliable time points to consistently measure behaviour. Results from the present study suggest that the 5M time point may provide the truest reflection on innate behaviour responses and could be the most beneficial data for GDUK to review.

GWAS has benefits above a pedigree-based analysis since GWAS identifies specific genome regions and candidate genes that are associated with traits of interest, which can provide a greater insight into the underlying biology of the trait. Results from GWAS indicated that Component 3 (Separation) at 8M demonstrated a suggestive level of significance for genomic-based analysis. The behaviours could be associated with a 'response to uncertainty', and there have been links reported between separation and anxieties in dogs (Amat *et al.*, 2020) and more recently for SNPs associated with fear responses in dogs (Bhowmik *et al.*, 2024). The top 20 most significant SNPs for each component were explored further and there were associations found to multiple psychiatric disorders in humans such as depression, Bipolar disorder, Schizophrenia, autism spectrum disorder, Alzheimer's disease and mood disorders and PTSD. The disorders or diseases can be associated with the anxiety and fear spectrum and Wormald *et al.* (2016) reported that aggression, compulsive disorders and phobias in dogs are linked to extreme anxiety. Separation behaviours can manifest in the dog to include vocalisations and destructive behaviours (Wang *et al.*, 2022) and it is thought up to 40% of dogs that attend behaviour clinics show behaviours associated with separation (Sherman, 2008), highlighting the prevalence within the dog population.

Rooney *et al.* (2016) suggested that selection, management and introduction to potential stressors could influence fearful and anxious

responses in puppies. GDUK could control some of these factors through appropriate training for puppy raisers on managing and introducing stressors, and where the behaviours are shown to be heritable, breeding selection may be possible. GDUK could breed from dogs with desirable traits and select puppies to continue through the guiding pathway based on the behaviours they have exhibited. However, the management of anxiety-related behaviours and introduction to stressors will be influenced by puppy raisers and the support they receive from GDUK. It could be suggested that the influence of environment is much greater than the influence of genetics, as demonstrated with a heritability of anxiety related behaviours at 5M ($h^2=0.29$), which reduced to 0.02 at 8M and 0.08 at 12M.

For many decades, GDUK have used the terms 'Confidence', 'Stress Resilience' and 'Suspicion' to define and attempt to measure behaviour (Evans, 2014; Asher *et al.*, 2016; Blythe, 2020). Arguably all of these terms could encompass a wider 'Anxiety' trait. Evans (2014) found the heritability of 'Confidence' using the crossbreed model was on average 0.04, 'Stress Resilience' had an average h^2 of 0.03 and 'Suspicion' an average h^2 of 0.05. Whilst these heritabilities are low, this is not surprising given the findings by (Asher *et al.*, 2016), demonstrating the large influence of staff members completing the report. The subjective terminology would have also added to variations seen in the phenotype and had more objective measures been used these heritabilities may have been higher.

During the behavioural assessment, the 'Surprising Events' task (associated with the startle response) was shown to have the greatest heritabilities, with over half of the 22 heritabilities above 0.4 being from the task. The task involved the use of three different items (a bin bag filled with crumpled paper, an umbrella and a metal sheet). Another task with several heritabilities over 0.4 was 'Novel Object'. The component comprising of interaction with a novel object was correlated to four of

the r-PWQ2 components including Separation and Anxiety. Similar tasks and items have been used by other studies to explore fear-related behaviours in dogs. The startle response using the opening of an umbrella (one of the items used in Surprising Events) and the presence of a novel toy were used by King *et al.* (2003), with PCA outputs grouping similar to the PCA grouping reported in the present study. Stellato *et al.* (2017) aimed to study fear responses in 31 companion dogs, and used a bin bag full of crumpled newspaper, also the same item used in Surprising Events in the present study, concluding that subtle behaviours were not useful indicators of fear responses in dogs. A recent study reported GWAS results for a startle response in a cohort of 615 dogs (Bhowmik *et al.*, 2024), with significant SNPs that were also identified in the present study. Results suggest that the startle response could be influenced by genetics, and thus potentially could be selected for by breeding.

7.5 Terminology conclusions

A topic that has remained in question throughout this thesis is what is understood by the term's 'behaviour', 'personality' and 'temperament', and what is the appropriate terminology to use when describing and discussing this topic. Did the behavioural observation assessment measure puppy cognition? Were the behavioural measures measuring personality or temperament?

Dog behavioural studies refer to all three terms within the literature, some using more the one term interchangeably. Studies refer to 'personality', e.g., Svartberg and Forkman (2002), Wiener and Haskell (2016) and Ilska *et al.* (2017). Others refer to 'temperament', e.g., Murphy (1998) and Duffy and Serpell (2012) and some refer to 'behaviour', e.g., Arata *et al.* (2010). There are then studies that use a combination of these terms. Hsu and Serpell (2003) refer to both 'behaviour' and 'temperament' Rayment *et al.* (2016) refer to 'behaviour'

and 'personality' and Jones and Gosling (2005) refer to both 'temperament' and 'personality'. It is clear to see why there is confusion within the literature and amongst researchers, especially when studies do not commonly refer to definitions for the terms used. Harvey (2014) proposed a schematic to demonstrate how the terms correlate for dog behaviour, indicating personality to be an overarching concept, with temperament and behaviour observations influencing personality (see Figure 1.2 in Chapter 1)

Within Chapter 1, the differing viewpoints and definitions for the three terms were discussed. During the development of the original Puppy Walker Questionnaire (PWQ), Harvey (2014) concluded that the PWQ measured 'personality' in guide dog puppies. Diener and Lucas (2019) define personality traits as consistent and stable. There were differences in the questionnaire groupings reported by Harvey (2014) during development and the r-PWQ published in 2020, and then again for the development of the r-PWQ2 (Chapter 2). Chapter 2 concluded that question groupings for all components were not the same for the three time points (5M, 8M and 12M), and therefore not all components were stable within this population. It could therefore be argued that the r-PWQ2 does not measure personality in the current study population, but rather behavioural characteristics that have been influenced by the dog's temperament (due to the genetic influences found). However, three components were stable between the time points and between the r-PWQ and r-PWQ2 (Anxiety, Body Sensitivity and Energy), suggesting these components to have the greatest validity within this guide dog population, and the components were shown to demonstrate pedigree-based heritabilities at 5M above 0.1. There are still great differences between researchers regarding the definition and assessment of personality within humans (Kernberg, 2016). It is therefore not surprising that there are differences in opinions for dog research.

Goldsmith *et al.* (1987) published “What is temperament”, stating temperament to be a group of related traits, and not a trait itself, with temperament noted to include elements such as ‘irritability, activity levels and fearfulness’. Cloninger (1994) developed the Temperament and Character Inventory (TCI) to determine ‘personality’ by measuring temperament and character in the human population. Temperament is the difference in associative learning between individuals for four dimensions whilst character reflects differences in values, goals, and self-conscious emotions (Garcia *et al.*, 2017). Cloninger suggests four dimensions of temperament; harm avoidance, novelty seeking, reward dependence and persistence. These domains are more applicable to what we can understand and know about animals and support the use of the term ‘temperament’ rather than personality to describe an animals’ higher behaviour. Garcia *et al.* (2017) further notes some elements of character to only exist in humans. Personality is arguably more applicable to humans than animals due to the inability to confirm the components of ‘character’ to be present in animals. Counteracting this viewpoint, studies that report personality to be present in animals was discussed in the Literature review in Chapter 1 (see section 1.2). With specific reference to the dog, Gosling *et al.* (2003) found four personality traits to be present in dogs, with Svartberg and Forkman (2002) and (Ley *et al.*, 2008) stating five underlying personality factors, although the five factors between these two studies do differ (see section 1.2.1 for details).

The TCI has been studied and used by numerous scientists since development. It was noted by De Fruyt *et al.* (2000) that the temperament dimensions stated in the TCI are “independently heritable” and “manifest early in development”. De Fruyt *et al.* (2000) compared the TCI to the well-known Five Factors Model of Personality, with considerable overlaps found between the two models. A study by Svrakic *et al.* (2002) utilised the TCI and defined temperament as “heritable neurobiological dispositions to early emotions (such as fear,

anger, and attachment), and their related automatic behaviour reactions (such as inhibition, activation, and maintenance of behaviour) in response to specific environmental stimuli (danger, novelty, and reward, respectively).”

Shiner *et al.* (2012) described advancements in the understanding of temperament, firstly, that not all temperament traits are stable early in life. Secondly, that emotional and cognitive processing are involved in temperament and thirdly, that temperament should be viewed as the result of biological and environmental factors interacting during development. Harvey (2014) defined temperament as “innate individual differences that have the ability to influence or predispose an individual in the development of specific personality traits”. Whilst Kernberg (2016) described that personality is made up of temperament, object relations, character, identity, ethical value systems, and cognitive capability. ‘Temperament’ being a subcategory to ‘personality’ is a contrast to the conclusion made by Harvey (2014), who suggested that personality is influenced by temperament. Diener and Lucas (2019) note that personality incorporates thoughts, feelings, and behaviours, and subsequently is an inappropriate term to use for non-human species where we are unable to determine feelings and thoughts, but more visible behaviours. There appears to be a circular argument where neither personality or temperament are fully defined or understood, with strong links to the two terms present in literature.

The behavioural observation assessment undertaken in the present study was based on an assessment developed by Bray *et al.* (2020), where the tasks attempted to measure a puppy’s cognitive abilities. The term ‘cognition’ has been discussed (Bayne *et al.*, 2019) and it was noted that no current definition appears to cover all uses of the term. Bayne *et al.* (2019) described how ‘cognition’ is a broad term that refers to how an animal processes information, whether by conscious or unconscious means, and how cognition is a popular term to add to any

form of problem solving. A consideration of mood and emotion could also be included when reviewing cognition. The assessment undertaken in Chapter 4 has been referred to as behavioural observation assessment rather than cognitive assessment. This is due to suggestion that cognition appears to have a multitude of subcategories, which appear to be undefined and not agreed within the literature. The tasks have also been referred to as ‘problem-solving tasks’ rather than cognition tasks.

7.5.1 Updated behaviour schematic

During this thesis, results and concepts suggest the schematic developed by Harvey (2014) does not fully account for the genetic influences on behaviour, and it could be argued that the term ‘personality’ is not appropriate for use within animals. The Harvey schematic is a contrast to the Cloninger (1994) schematic, where Cloninger refers to personality as the higher overarching principle, while Harvey refers to temperament as the higher overarching principle. The definition for temperament suggested by Svrakic *et al.* (2002) appears to be consistent with the findings of the present study, and due to anthropomorphic issues present using the term ‘personality’, the author of this thesis proposes that this be removed from the Cloninger and Harvey schematics when describing behaviour in dogs.

New use of terminology is suggested following the findings from this thesis. It is proposed that ‘temperament’ is the overarching principle for behaviours seen in dogs, being influenced by genetics and the environment (Rothbart *et al.*, 2000). A new proposed schematic (see Fig 7.2) is an inversion of Harvey’s (Harvey 2014) with behavioural observations appearing at the top of the schematic and highlighting the influences of assessors, the environment and genetics throughout all stages. Within the new schematic, an observation is defined as “the action or an act of observing, watching, or noticing” (Dictionary, 2024b).

Behavioural observations can be seen and described for individual animals, and these can be measured by different methods (e.g. questionnaires, behavioural assessments/tests), which is demonstrated on the schematic. Questionnaires consist of items and scales to measure behaviour, whereas assessments measure factors or tasks. The importance of reliability, consistency and validity for both the data types is highlighted on the schematic.

Harvey (2014) refers to the term 'dimension' to be a subcategory of personality. The definition for dimension is 'measurable or spatial extent of any kind' or 'magnitude, extent, degree' (Dictionary, 2024a) and the term is defined by Harvey as "higher order trait constructs (commonly identified by multi-level cluster or factor analysis) that summarise the shared variance of the correlated traits they comprise". 'Dimensions' are formed by 'traits' which are referred to as "a collection of correlated behavioural responses identified by statistical evaluation (e.g. factor or principal component analysis)" so the use of the two terms is unclear and both refer to factor analysis. Harvey (2014) generally refers to traits throughout the thesis, and there are some conflicting uses of terms, such as one table referring to 'distractions' as a dimension, and then 'distractibility' is referred to as a 'scale'. It could be argued that one term would be sufficient. Blythe (2020) then refers to Distraction as a 'personality characteristic' and a 'factor', adding further confusion as to the most appropriate term to use.

Consequently, the new proposed schematic refers to behavioural components which infer temperament. Behavioural components are comprised of behavioural observations which are also impacted by experiences, situations/events and exposure. The present study has referred to components extracted by PCA during multiple studies. It is suggested that by removing a level and reducing the terms used, this will be simpler and clearer for readers to follow. It will not always be

appropriate to have a higher category, but should there be a need, these could be referred to as a group of components.

Using examples from the data collected during the present thesis highlights how the schematic works. Chapter 2 collected data from a behavioural questionnaire (the r-PWQ2) through a set of 39 individual items. Following a PCA, ten components were extracted at each point which were then given a label and are suggestive of an individual's underlying temperament due to heritability findings in Chapters 3 and 6. The data collected using the r-PWQ2 was shown to be influenced by several factors. Similarly, data collected during Chapter 4 for the behavioural observation assessment collected observations for different behaviours depending on the task undertaken and explored aspects of problem solving ability. Assessor influence was shown to be present in Chapter 5 due to inter and intra-reliability differences. Some behaviours were shown to have pedigree-based heritabilities and the behaviours grouped to form higher components e.g. interaction with novel item. These behaviours were also then correlated to questionnaire components, showing the links between the data to inform an individual's temperament.

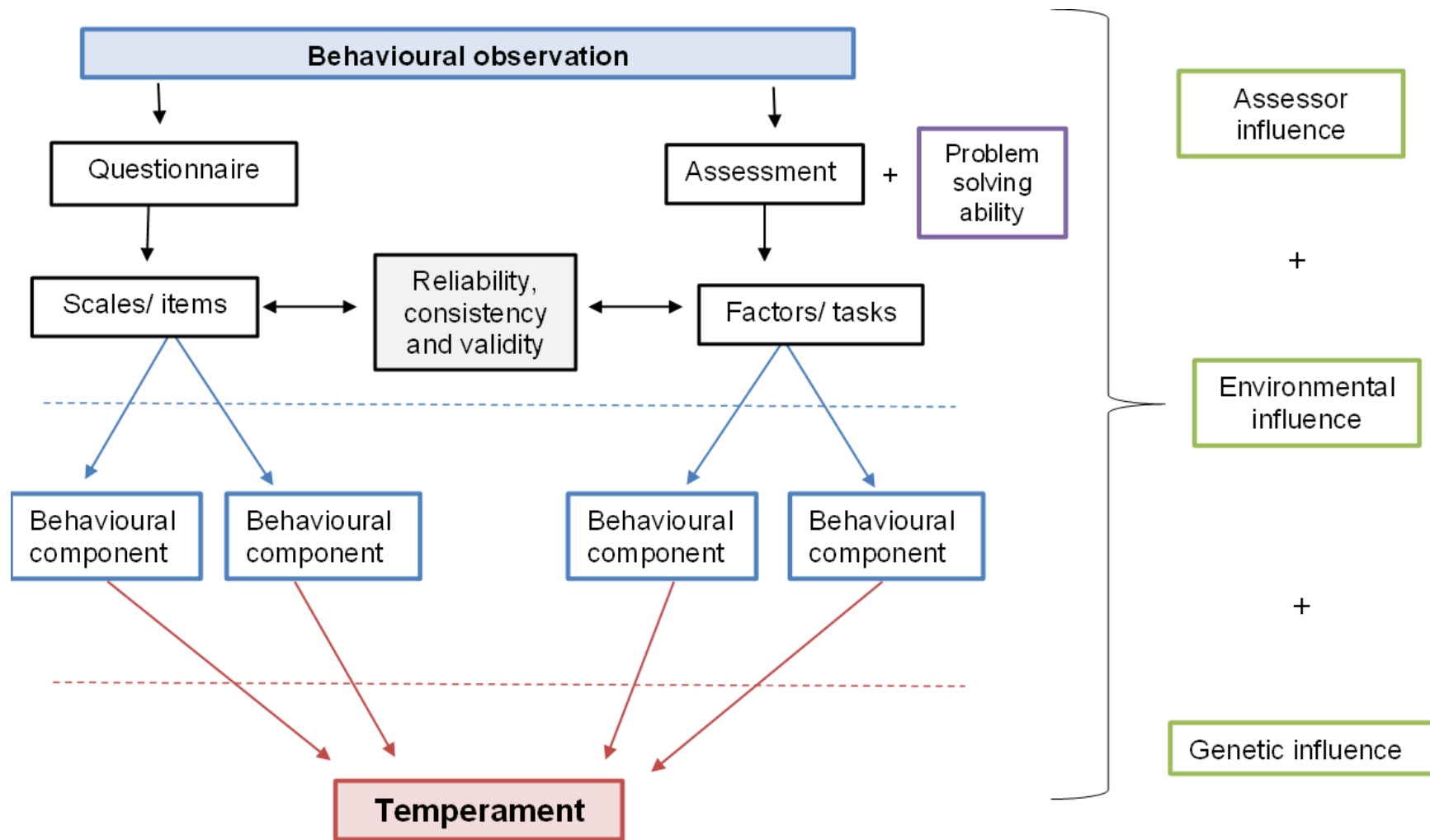


Figure 7.1 Updated schematic for temperament and behaviour.

7.6 Selective breeding for behaviour

Dogs are regularly referred to as ‘a man’s best friend’, but as suggested by Diederich (2024), is man the best friend of dogs? Dogs have been bred and selected over many years to meet the requirements of humans, whether for working roles, as companions/pets or to profit from breeding. There are ethical concerns related to dog breeding, such as, overbreeding e.g. puppy farming, selective breeding for appearance and whether dogs should be bred to benefit humans e.g. assistance dogs.

When reviewing breed popularity, Ghirlanda *et al.* (2013) reported the most popular dog breeds were not those with the most desirable behaviour. The study found correlations between popularity and fear/anxiety related behaviours (including separation problems, fear of other dogs, and aggression directed toward the owner). These findings highlight that breeding for behaviour may not be prominent within dog breeding, but also that social influences have influenced dog breed popularity, and thus the demand for breeding specific breeds.

Ethical concerns related to ‘fashionable’ dog breeding have been raised by researchers, where appearance has been used as the primary selection tool. Selective breeding in this way has been suggested to increase behavioural problems such as anxiety and aggression, whilst also impacting the health and wellbeing of dogs (Menor-Campos, 2024; Morel *et al.*, 2024). Selecting dogs in this way also influences the breed population, increasing the presence of inbreeding and a loss of genetic diversity (Rossi, 2013); with modern breeds distinguished genetically by aesthetic traits (Morrill *et al.*, 2022). Breeding for one specific trait (whether that be appearance, coat colour etc.) may result in an increase in the prevalence of genetic disorders (Axelsson *et al.*, 2021). For example, the selection for coat colour in Dalmatians has been associated with deafness in the breed (Cline, 2012). Scientists are striving to raise awareness and outline genetic problems in

popular breeds of dogs to highlight risks to prospective pet owners (UFAW, 2024).

Contrasting views and conflict can be seen between dog breeders, associations such as the Kennel Club and veterinary professionals (Wanner, 2017). It is clear to see why there may be some confusion between the groups associated with dog breeding. Whilst promoting responsible breeding and health testing, kennel clubs also promote that appearance is important due to animals being judged by physical appearance standards, and as such the organisations are not helping to promote the prioritisation of breeding healthy dogs (Menor-Campos, 2024); unless of course these standards are revised to prioritise health. It has been suggested that in more recent years the influence of kennel clubs is reducing, with more dogs being sold without a pedigree (Proschowsky *et al.*, 2025). Whether distinct breeds of dogs should exist has been debated (Hens, 2009) and if a group of healthy mixed or cross breeds would be more appropriate. This would go against the views of kennel clubs, where distinct breeds and standards are promoted. In more recent years cross breeds of dogs have become more popular (Burnett *et al.*, 2022), where cross breeds are thought to be more ‘healthy’ due to the presence of hybrid vigour and heterosis (Proschowsky *et al.*, 2025). A recent study of three popular cross breeds (Cavapoo, Cockapoo, and Labradoodle) were compared to the parent pure breeds, with limited differences in health incidence found seen (Bryson *et al.*, 2024). More popular dog breeds have also been associated with a greater number of inherited disorders (Ghirlanda *et al.*, 2013), yet this does not seem to stop the demand.

Following debates on ethical dog breeding, it is not surprising that there has been an increase in awareness of assistance dogs welfare (Salmon *et al.*, 2022), and whether it is appropriate to use dogs in different situations (Foltin and Glenk, 2023). Taş (2025) challenged the assumption that all assistance dogs ‘enjoy’ their role and suggested that there is a need to pay attention to the individuals’ behavioural cues. In contrast, Gibson and Oliva (2022) reported that public perceptions of assistance dogs were that they are

'happier' than companion dogs, due to assistance dogs being in constant company with their handler. Assistance dog organisations are under the spotlight, with a need to ensure ethical standards are upheld, the number of dogs bred is appropriate, that welfare is maximised, that dogs display appropriate behaviours and that qualification rates are high to reduce the number of dogs needing to be rehomed as pet dogs.

The moral acceptability of selective breeding in dogs has been discussed (Bovenkerk and Nijland, 2017). The findings from the present study suggest it could be possible to selectively breed for behavioural displays within this population, with the greatest potential for selection to be for anxiety-related behaviours (response to uncertainty). Should similar findings be present in other populations, and if the reaction to uncertainty were more predictable in dogs, appropriate handling and training advice could be provided to owners to increase desirable behaviours. Breeding away from these behaviours for the wider dog population could help to reduce the prevalence of dogs relinquished to shelters each year, reduce the need of behavioural interventions and ultimately improve animal welfare. In addition, there could be a positive influence on human health and safety. The incidence of dog bites in the UK is reported to be 18.7 per 1000 people (Westgarth *et al.*, 2018), with dog bites linked to anxious behaviours (Reisner, 2003). This may be reduced should there be less anxious dogs.

7.7 Wider implications

Results from the present study suggest there may be wider applications to the field of dog behaviour and potentially other species. The behaviours displayed (specifically anxiety associated behaviours) in this population were found to be heritable, with genetics influencing a proportion of behavioural variation seen. The breeds used in the present study (Labradors and Golden Retrievers) are also popular pet dog breeds (Forsyth *et al.*, 2023). Success has been shown when selecting against anxiety-related behaviours in rats; a twofold difference was seen after eight generations and maternal interactions

were shown to have a small influence on the behaviour (Stead *et al.*, 2006). The study suggests improving anxiety-related behaviours may be possible in animals, but whether this is applicable to other species is unknown, and the time taken to reach eight generations would vary considerably by species. Should the findings in the present study also be present in other populations or breeds of dogs, there is the potential to promote breeding less anxious dogs.

DNA testing for inherited health conditions are widely used in dog breeding, with a range of tests available depending on the breed of dog (Meyers-Wallen, 2003). In the future, when there is a more joined up approach to describing and reporting behaviour (as proposed in section 7.5), it may be possible to explore the development of DNA testing/selection for behaviour in dogs. Yet for this to be possible, researchers would need to work together and collect similar phenotypes from a wide number of dogs and breeds. If it is possible to select and breed for specific behaviours, there will be the need to breed less dogs and the potential to select dogs suitable for certain roles.

Dogs are known for their acute sense of smell and ability to detect a multitude of different scents (Horowitz *et al.*, 2013). The present study found one measure from the Odour Discrimination task to have a heritability of 0.54 in young puppies, suggesting that engagement with food items may have some genetic influence in this population which can be seen at a young age. A behavioural assessment for juvenile guide dog puppies also revealed one item to be associated with training outcome at five and eight months of age; orientation towards food (Harvey *et al.*, 2016). Genetic links to dog obesity have been found in Labrador Retrievers (Raffan *et al.*, 2016), including dogs from a guide dog population. Dogs are also commonly trained with reward based methods such as food (Hiby *et al.*, 2004). A balance between breeding dogs that are food motivated (to support with training) but also healthy and not prone to obesity is a fine balance. Links between canine and owner obesity have been reported, with overweight owners being three times more likely to have overweight dogs (Suarez *et al.*, 2022). Genes were

identified in Chapter 6 that relate to eating disorders in humans, a form of food orientation and studies have found links to temperament traits and eating disorders in humans (Fassino *et al.*, 2002). These findings may strengthen our understanding of the links between psychological states and feeding behaviours in humans through data collected on the dog. There are also implications for human health as obesity prevalence is rising, with an estimate of 25% of the population being obese, and the potential for this to rise to 31% (Janssen *et al.*, 2020). Understanding behaviours associated with feeding behaviour could influence obesity in both the dog and humans.

Profiling and selection of dogs for specific working roles, e.g. scent detection or police work, may be possible should heritabilities for specific behaviours be present. If organisations could identify the most suitable career for a dog from a young age, resources could be saved, and dog welfare maximised. A similar initiative was trialled by GDUK, where ‘purple’ profiled puppies from PPA (Asher *et al.*, 2013) were placed in alternative assistance dog programmes. This could be applied more broadly, but would require organisations to work collaboratively together, which can be challenging when charities/organisations are competing for funding and resources e.g. volunteers. It may also be possible to utilise selection for other behaviours and situations. For example, in the present study Impulsivity had a heritability of 0.08 at 8M and Body Sensitivity a heritability of 0.10 at 5M. More ‘impulsive’ dogs may be more appropriate for police work or scent work, and dogs with body sensitivity would not be suitable for the guiding role where a dog is required to wear a harness. Dogs have been able to detect positive Covid (SARS-CoV-2) samples with 83% sensitivity and 96% specificity when trained for only one week (Jendrny *et al.*, 2020), demonstrating the success of scent detection dogs. Should it be possible to select for, or breed towards, dogs with an acute sense of smell, there may be wider benefits to humans if more scent detection dogs were available. For companion/pet dogs, understanding the heritability of different behaviours could aid breeders with selling puppies to the most appropriate home.

Citizen science initiatives exist worldwide in a range of different situations and involve members of the public participating in research projects (Shirk *et al.*, 2012). Following the findings from this thesis and the suggestion that it may be possible to profile or select dogs, the combined use of ML and citizen science could help to broaden our understanding and support selection in the future. AI (2025) has successfully developed live data tracking for farm animals to highlight welfare concerns in real time. This type of technology and reviewing data live for individuals has wider applications to other species and settings e.g. rehoming shelters, breeding facilities, zoological parks and wild animals. The results from Chapter 5 suggest it would be possible to use ML to measure puppy behaviour. Due to the importance of early behavioural indicators, and the higher heritabilities seen for young puppies in Chapter 4, it may be possible to develop software to measure and monitor wider puppy behaviour. Should researchers be able to determine a small battery of tasks using everyday household items, owners /breeders could film individual puppies (and their parents, siblings, and any future offspring), upload the data onto a website/ platform and ML be used to extract data and make suggestions for how best to support the individual and maximise welfare.

7.8 Study limitations

The main limitations of the present study were that the cohorts of dogs and the time point used for behavioural data were not the same between the Chapters (Table 7.1). There were links between some studies, but not all data were collected from the same individuals due to study design. Different populations were used rather than following a cohort of dogs through their guide dog training journey. This journey follows puppies as they are raised with volunteers from eight weeks of age until around 12-14months of age when they enter formal training. If the dog is to succeed in formal training, a client will then train with the dog until the partnership 'qualify'. A dog can be withdrawn from the process at any point or stage if they are deemed unsuitable based on health or behaviour.

Table 7.1 Summary of data, population size and analysis by Chapter

Chapter number	Data type	Age of subjects	Population size	Analysis
Chapter 2	r-PWQ and r-PWQ2	5M, 8M, 12M	551, 2137, 1537,	PCA
Chapter 3	r-PWQ2	5M, 8M, 12M	492, 2063, 1375	Pedigree-based heritability
Chapter 4	Practical assessment	6-10 weeks	399	Pedigree-based heritability
Chapter 5	Practical assessment	7-8 weeks	25 and 80	Reliability and machine learning
Chapter 6	r-PWQ2	8M and 12M	96	Genomic-based heritability

The validity of the r-PWQ2 remains in question as the data used between the three time points was not consistent, influencing the content validity. The r-PWQ2 was however significantly correlated to the r-PWQ when comparing the two data questionnaire structures, and the r-PWQ demonstrated validity during development. Further work on the criterion validity could be undertaken to associate the r-PWQ2 to the practical assessment undertaken in Chapter 4. Validity was demonstrated for three components (Anxiety, Body Sensitivity and Energy), with the same groupings found between r-PWQ and r-PWQ2 and for the three time points, suggesting these components to be the most reliable between the populations. Interestingly, these were also the components that demonstrated higher pedigree-based heritabilities at 5M in Chapter 3 and links to the behavioural assessment groupings in Chapter 4. Some of the differences made to the r-PWQ2 could be due to the different groups of dogs present in the 5M, 8M and 12M data, which could be overcome by undertaking a cohort approach in the future. GDUK collects the r-PWQ as business as usual for all puppies in puppy raising. It would therefore be suggested that the r-PWQ groupings be explored in more depth in the future to ensure valid data is used for heritability analyses.

Power calculations were not made to determine sample size as the data collection was limited for each chapter by practical constraints and resource limitations. For the questionnaires, sample size was dictated by questionnaires available prior to a specific date. Unfortunately, data numbers

were impacted by the Covid-19 pandemic. Data collected on puppies raised during the pandemic were not analysed, and therefore data collection ended earlier than planned. Therefore, only around 500 questionnaires were available for the 5M questionnaire analyses. A larger sample would have been preferable. However the sample size is greater than the sample used to develop the PWQ (Harvey, 2014). The number of dogs genotyped was dictated by available funds. The number of puppies that undertook the behavioural observation assessment was based on the available puppies within a calendar year. Data for the ML pilot was limited by resources and availability of time to clip and trim videos. Using mean values reported by Bray *et al.*, (2020), power calculations suggest that with a sample of 399, using a 95% confidence level and 5% margin, 100% power could be detected, suggesting the sample size to be adequate in Chapter 4 (ClinCalc, 2025). The number of questionnaires collected in Chapter 2 was considerably larger than the population used to develop the questionnaire, suggesting this sample size to be sufficient. For the pedigree-based heritabilities, between 500 and 2000 dogs were reviewed. To detect a h^2 of 0.4 (the suggested heritability for behavioural traits; Ilska *et al.*, 2017), 500 dogs would provide a power of 10% and 2000 dogs a power of 72%, suggesting a sample over 2000 would be advisable (Visscher *et al.*, 2014). The genomic-based heritability sample size of 96 was therefore considerably smaller than required to detect significant values.

A limitation of some heritability models undertaken in Chapter 3 (5M r-PWQ2) and Chapter 4 (behavioural assessment) was that data was collected for a cohort sample representing one year of puppy production. The relationship matrix used for heritability analyses will be smaller than for data spanning multiple years. Breeding dams are likely to have only had one litter within the period for data collection, and only full sibling and half sibling relationships will have been analysed. If data were collected for multiple years, parent and offspring relationships could be studied. This may have accounted for some of the difference seen in heritabilities between the 5M sample and the 8M and 12M sample. Additionally, the analysis did not

separate pure breeds and cross breeds as undertaken by (Evans, 2014), an area which could be explored in the future with a larger sample size.

Another limitation was that not all breeds owned by GDUK were investigated. German Shepherd Dogs (GSD) are bred by the GDUK breeding programme, but in much smaller numbers than the more common Labrador and Golden Retrievers. In 2023, GDUK produced 64 pure-bred GSD puppies and 59 cross-bred GSD puppies born alive, accounting for 9.9% of the puppies produced that year. GDUK also produces a small number of Labradoodles (Standard Poodle x Labrador; $n=13$, 1.1%). Both the GSD, GSD crosses and Labradoodles were removed from analyses due to the smaller numbers, and the differing behaviour characteristics present in the breeds. For applications to the wider dog population, the present study included a limited number of breeds within a highly managed population, therefore how pertinent the findings would be to other populations and dog breeds is unknown.

Harvey (2014) found the inter-rater reliability of the Puppy Walker Questionnaire (PWQ) when compared to a questionnaire for Puppy Development Advisors to display significant correlations at 8M and 12M. There was a distinct lack of agreement between volunteers and staff in reports of dogs at 5M, suggested to be due to the short amount of time staff will have seen puppies by the 5M point. There were differences found in the component structure between the present study population and the one reported by Harvey (2014), hence the r-PWQ2 was formed. The present study could have been improved had the inter- and intra-rater reliability of the r-PWQ2 been undertaken. Chapter 5 has highlighted the importance of reviewing the inter- and intra-rater reliability of practical assessments.

The present study has demonstrated the great influence of puppy raisers and the environment on puppy behaviour. As such, assistance and working dog organisations should review and analyse training modules for puppy raisers, specifically focusing on less heritable therefore those less likely to be influenced by breeding. Chapter 2 reported the impact of puppy raiser

experience (defined as the number of puppies previously raised by the individual) on behavioural measures, and it was therefore planned for the heritability analysis to include puppy raiser as a random effect. However, when the model was run with the puppy raiser included, heritabilities were inflated and following further investigation it became clear that puppy raisers commonly had puppies from the same sire/dam when they had raised more than one puppy. Therefore, puppy the raiser ID was not randomised and was confounded with dog genetic relationships. A larger sample size would allow for the inclusion of more variables within the heritability models, and it may be possible to include puppy raiser effects if puppies were randomly allocated to volunteers.

7.9 Future work at GDUK

Whilst the main aims of the study were met during this thesis, results have highlighted more questions and possibilities for future work. The outcomes from Chapter 3 indicated that the 5M time point showed the highest pedigree-based heritabilities when compared to the 8M and 12M questionnaires. The outcomes of this study found that the genomic-based heritabilities were higher than pedigree-based heritabilities. Undertaking a larger genomic-based heritability analysis on the 5M data is suggested because a SNP-based analysis would allow for a wider exploration of genomic correlations and the ability to undertake more GWAS rather than basing results on phenotype for related individuals. GDUK are routinely collecting DNA swabs from puppies born 2020 onwards (Guide Dogs, 2020b) as part of the 'Born to Guide' project. The resulting genotype data could be used to analyse the r-PWQ2 in the future, and for a much larger sample and for dogs that have all three behaviour questionnaires completed. Additionally, a larger sample would enable analysis by breeds and cross breeds, as a clear distinction was present for pure and cross breeds when reviewing the population stratification for the 96 genotyped dogs. A larger sample size would also equate to more generations of dogs (and more relationship

connections) being present in the data set, which could impact the heritabilities studied.

As it may take years to source the genotype data and r-PWQ2 for a large sample of dogs, GDUK could consider creating Estimated Breeding Values (EBVs) for the 5M behavioural components that are suggested to show heritabilities above 0.1. Behaviour accounts for most of the withdrawals from the guide dog programme. Were EBVs for behavioural selection to be implemented by GDUKs, the percentage and number of behaviour withdrawals could be reviewed over to time to determine whether the selection had an effect on qualification outcomes.

Similarly, the puppies in Chapter 4 will have had DNA collected and in future reviewing genomic-based heritability estimates of the behavioural observation assessment would enable a deeper understanding of the early innate influences on behaviour. Higher heritabilities ($h^2=0.4$ or above) were reported for the behavioural observation assessment data when compared to the r-PWQ2. Following the trend found for results using the r-PWQ data, it would be interesting to see how heritabilities would compare for genomic-based estimates using the behavioural observation data. A recent study found measures associated with cognitive abilities were not associated with working success in a cohort of Australian guide dogs (Foraita *et al.*, 2024). The links to outcome as guide dog and the behavioural observation assessment have not been explored as part of this thesis due to the timeframe for data collection and the time it takes for dogs to reach outcome as a guide dog (around approximately two years of age). It is suggested that GDUK follow up on the outcome of the 399 puppies assessed in Chapter 4 and determine if the heritable behaviours show any correlation to success as a guide dog.

Findings presented in Chapter 5 highlight the importance of reliable and accurate data. GDUK need to continually review data to ensure it remains relevant in different cohorts of dogs. This study found differences in the

original PWQ developed by Harvey (2014), therefore the r-PWQ was created. The main differences were a reduction in the number of questions between the r-PWQ and the r-PWQ2 and the formation of 10 components in the r-PWQ2 compared to nine in the r-PWQ. The questionnaire groupings and heritabilities should be reanalysed at intervals, suggested to be every five years. This is due to the average breeding dog career length being approximately five years, so repeating analysis will be representative of different breeding animals. Future work could include the use of ML to objectively analyse dog behaviour. Results from Chapter 5 indicated that for a small sample, it was possible to use ML to measure the time a puppy spent interacting with an item. This could be applied more widely throughout the GDUK organisation. For example, CCTV videos collected of puppies between seven and eight weeks of age when at the GDNC could be used to analyse puppy interactions with siblings and items placed within the kennel. This type of analysis would provide real time accurate data with little staff input. The data could be analysed by ML almost instantly if videos could be uploaded to a system. Data such as time spent interacting with a specific item/handler, time playing with littermates or duration of time spent sleeping could be captured and reviewed to aid with volunteer placement. Due to the continued influence of the puppy raiser and environment on behaviour, should GDUK be able to place puppies in the most appropriate home, this could influence future behaviours and potentially the number of dogs suitable to enter training.

7.10 Conclusion

This study met its aims by successfully reviewing data collected through questionnaires and practical assessments for a cohort of guide dog puppies. The heritabilities of measures were reviewed alongside reliability, and the use of genotype data and machine learning to provide a range of perspectives and a deeper understanding into the genetic influences on guide dog behaviour. Results indicate that behaviours associated with guide dog success have a heritable and genetic basis that the GDUK breeding

programme could explore further. It would be possible to calculate behaviour estimated breeding values for selected r-PWQ2 components at 5M of age, which also suggests there would be the potential for the organisation to use genomic breeding values in the future. There is the potential to influence guide dog behaviour in the future and therefore provide more dogs to aid people with a visual impairment.

The present study has reviewed the terminology used when describing dog behaviour and concluded that the term 'temperament' should be used when describing a group of behavioural components due to the influence of genetics on dog behaviour found during this study and by others. An updated schematic has been proposed and examples given to demonstrate how the schematic works for the data collected during the present thesis.

This study has wider impacts to understanding the genetic influences on dog behaviour, suggesting anxiety related, excitement and separation behaviours may all be influenced by breeding. There could be the potential to breed for more desirable behaviours in dogs, improving welfare and influencing owner-dog relationships. Findings from the Genome Wide Association Study (GWAS) demonstrated genes approaching significance that were also linked to behavioural phenotypes in humans, such as attention deficit hyperactivity disorder (ADHD), alcoholism, depression and schizophrenia. Results also suggested that Excitement and Separation behaviours seen in guide dog puppies could be influenced by sensory perception and olfaction, which adds to the understanding of the underlying biology affecting dog behaviour. Using the dog as a model species for behavioural disorders could widen our understanding of conditions in humans and other animals, advancing findings further in this field.

Chapter 8. Appendices

8.1 r-PWQ email and survey introduction

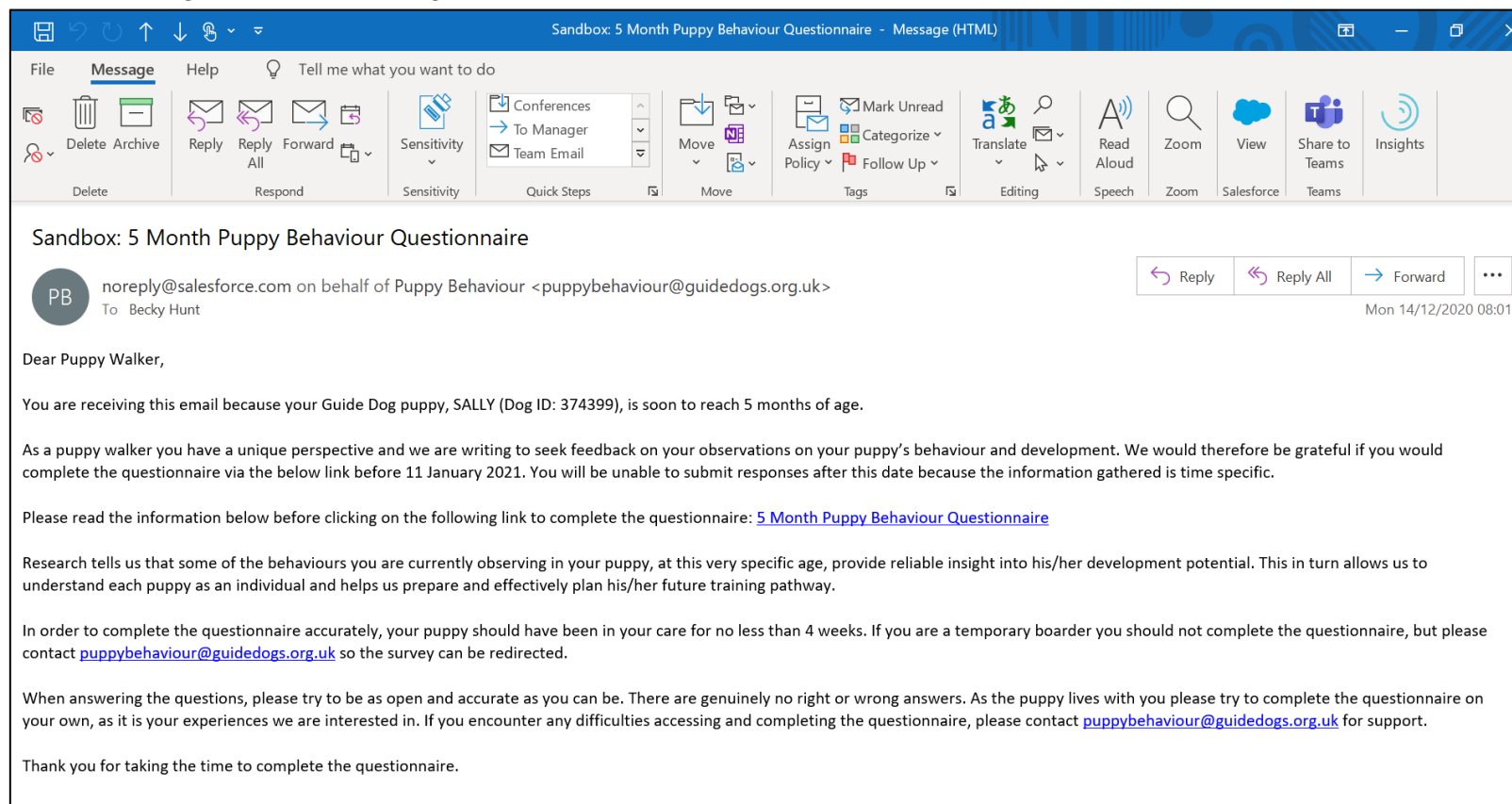


Figure 8.1 r-PWQ2 email template sent to a Guide Dogs UK puppy raiser inviting participation in questionnaire completion



Puppy Behaviour Questionnaire - 8 Months

Contact & Dog Details

Welcome to the Puppy Walker Questionnaire!

This questionnaire will take approximately 20 minutes to complete. Please bear in mind that there is no wrong or negative answer to any of the questions; these are all normal behaviours that can be seen in any puppy or even adult dog. Also bear in mind that they are still puppies after all! So no matter how you describe your dog please rest assured that no answer is a bad one. Just try and answer as honestly as possible. We very much appreciate your help.

First, please check that your name, your puppy's name and puppy ID are correct. If any are incorrect or you have any other questions please contact your Puppy Training Supervisor.

****Please only complete this survey if you are this puppy's volunteer puppy walker and your puppy has been in your care for a minimum of four weeks.****

1. Contact name *

2. Puppy name *

3. Puppy identification number (e.g. 360202) *

Next

Figure 8.2 r-PWQ2 questionnaire opening page.

8.2 r-PWQ questionnaire

Table 8.1 r-PWQ 2 question component group, question number (N) and wording, including the direction of association between the individual item within the group (negatively valanced items were transformed before average scale scores were calculated).

Component group	N	Question wording	Direction within group
r-Separation - related behaviour	1	Appears restless/agitated or paces when left, or about to be left	+
	2	Whines when left, or about to be left	+
	3	Barks when left, or about to be left	+
	4	Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left	+
	5	Loses its appetite when left, or about to be left	+
	6	Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	+
r-Excitability	7	Exhibits a high degree of excitement (jumps up; barks; coughs etc.) just before being taken for a walk	+
	8	Exhibits a high degree of excitement (jumps up; barks; coughs etc.) just before being taken on a car trip	+
	9	Exhibits a high degree of excitement (jumps up; barks; coughs etc.) when visitors arrive at your home	+
	10	Is excessive, difficult to control and if it lunges is hard to hold back	+
	11	Is hyperactive, restless, has trouble settling down	+
	12	Barks persistently when alarmed or excited	+
	13	Is self-controlled and calm <i>[appears also in Trainability]</i>	-
r-Attachment Attention Seeking	14	Tends to follow you (or other member of household) about the house from room to room	+
	15	Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	+
	16	Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person	+
	17	Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal	+
	18	Displays a strong attachment for one particular member of the household	+
Energy Level	19	Returns directly to you if startled or frightened	+
	20	Is active and energetic	+
	21	Is playful	+

General Anxiety	22	Is obviously disturbed by loud or unexpected sounds	+
	23	Is spooked by odd or unexpected things or objects	+
	24	Is anxious or uneasy in new situations	+
	25	Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)	+
Animal Chase	26	Chases birds or squirrels (or would like to)	+
	27	Chases cats (or would like to)	+
Body Sensitivity	28	Is uneasy with being physically handled/groomed	+
	29	Attempts to move away when you start to groom it	+
Distractibility	30	Pulls (including lunging) towards unfamiliar dogs	+
	31	Is motivated towards/distracted by food on the ground and or on tables/shelves	+
	32	Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	+
	33	Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	+
Trainability	34	Attention can be attracted easily but it loses interest soon	-
	35	Will look at you when you talk to it directly in the home environment	+
	36	Attention can be easily distracted	-
		Is self-controlled and calm [<i>appears also in Excitability</i>]	+
	37	Needs obedience commands repeating to get a response	-
	38	Seems like it doesn't listen even if it knows someone is speaking to it	-
	39	Stay's/Wait's' when instructed to	+

Table 8.2 Additional information questions collected on puppy raising environment and puppy routine.

N	Question wording	Scoring options
40	How many people are permanent residents in your house	Number of males and females: Adults over 18; children aged 13-18; children aged 8-12; children aged 4-7; children aged 3 and under.
41	How many other animals are resident in your house	Number of: Dog; Cat; Rabbit; Chicken; Other.
42	Please describe the location of your house	Urban; Suburban; Rural.
43	Did your puppy attend puppy classes	Yes, No.

44	Was your puppy taken into the workplace	Yes, No.
45	Over the last three months has your puppy shown any of the following behaviours	Chew objects they are not supposed to; Carrying objects they are not supposed to; Mouthing people; Nudge nuzzle or paws when you are eating; Barking when people come to the door; Jumping up at people; Showing possessive behaviour over toys or food (e.g. growling when you approach); Urinating or defecating indoors; Chases other animals; Scavenging or stealing.
46	During training is your puppy motivated by	Vocal praise; Food reward; Physical touch (e.g. pat); Play/toys.
47	Please tick the sentence below which best describes your puppy's spending behaviour	Only ever spends in an allocated spending area, never spends in inappropriate places; Occasionally spends outside of their allocated spending areas, in an inappropriate place; Frequently spends outside of their allocated spending area, in an inappropriate place; Always spends outside of their allocated spending areas, in an inappropriate place.
48	Please tick the sentence below which best describes your puppy's defecation frequency	The dog defecates either once, twice or three times daily, but the frequency is strictly reliable (i.e. the dog defecates the same number of times per day); The dog defecates either once, twice or three times daily, but the frequency can vary; The dog defecates four or more times daily; The dog's defecation frequency is very erratic/variable.
49	How often your dog has been coprophagic in the last three months during a free run	Never – Always (0-100) sliding bar
50	How often your dog has been coprophagic in the last three months while on a lead	Never – Always (0-100) sliding bar
51	How often your dog has been coprophagic in the last three months in the garden	Never – Always (0-100) sliding bar
52	How often your dog has been coprophagic in the last three months in any other situations	Never – Always (0-100) sliding bar
53	Please indicate the nature of the coprophagic behaviour	Eats own faeces; Eats faeces of other dogs; Eats faeces of other animals (e.g. sheep or cows).

8.3 PCA extraction results

8.3.1 5M PCA extraction results

Table 8.3 5M PCA extraction for r-PWQ2 questions

Question	Extraction
Appears restless/agitated or paces when left, or about to be left	0.71
Whines when left, or about to be left	0.74
Barks when left, or about to be left	0.67
Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left	0.43
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	0.55
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk	0.78
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip	0.78
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	0.56
Is excessive, difficult to control and if it lunges is hard to hold back	0.67
Is hyperactive, restless, has trouble settling down	0.59
Barks persistently when alarmed or excited	0.52
Is self-controlled and calm	0.50
Is obviously disturbed by loud or unexpected sounds	0.73
Is spooked by odd or unexpected things or objects	0.80
Is anxious or uneasy in new situations	0.64
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)	0.67
Chases birds or squirrels (or would like to)	0.71
Chases cats (or would like to)	0.71
Pulls (including lunging) towards unfamiliar dogs	0.70
Is motivated towards/distracted by food on the ground and/or on tables/shelves	0.55
Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	0.67
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	0.74
Attention can be attracted easily but it loses interest soon	0.42
Will look at you when you talk to it directly in the home environment	0.66

Attention can be easily distracted	0.47
Needs obedience commands repeating to get a response	0.53
Seems like it doesn't listen even if it knows someone is speaking to it	0.57
Stays/waits when instructed to	0.45
Is uneasy with being physically handled/groomed	0.77
Attempts to move away when you start to groom it	0.77
Tends to follow you (or other member of household) about the house from room to room	0.54
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	0.51
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person	0.67
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal	0.58
Displays a strong attachment for one particular member of the household	0.45
Is active and energetic	0.80
Is playful	0.83

Table 8.4 Correlation matrix for 5M PCA using Varimax rotation. Component loadings above 0.3 (whether positive or negative) are shown and those below 0.4 shown in red. Questions that appear in two components are highlighted in bold where both loadings are above 0.4.

Question	Component									
	1	2	3	4	5	6	7	8	9	10
Appears restless/agitated or paces when left, or about to be left		0.82								
Whines when left, or about to be left		0.85								
Barks when left, or about to be left		0.80								
Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left		0.59								
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone		0.67								
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk								0.84		
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip								0.85		
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	0.34			0.57						
Is excessive, difficult to control and if it lunges is hard to hold back	0.32			0.68						
Is hyperactive, restless, has trouble settling down				0.57					0.37	
Barks persistently when alarmed or excited				0.57	0.30					
Is self-controlled and calm				-0.44						0.44
Is obviously disturbed by loud or unexpected sounds			0.83							
Is spooked by odd or unexpected things or objects			0.85							
Is anxious or uneasy in new situations			0.73							
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)			0.78							
Chases birds or squirrels (or would like to)						0.81				
Chases cats (or would like to)						0.79				
Pulls (including lunging) towards unfamiliar dogs	0.52			0.46		0.44				
Is motivated towards/distracted by food on the ground and/or on tables/shelves	0.48					0.46				

Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	0.63			0.35						
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	0.62			0.34		0.43				
Attention can be attracted easily but it loses interest soon	0.62									
Attention can be easily distracted										0.77
Needs obedience commands repeating to get a response	0.66									
Seems like it doesn't listen even if it knows someone is speaking to it	0.58									-0.35
Stays/waits when instructed to	0.45									-0.51
Will look at you when you talk to it directly in the home environment										0.57
Is uneasy with being physically handled/groomed									0.84	
Attempts to move away when you start to groom it									0.82	
Tends to follow you (or other member of household) about the house from room to room					0.55					
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	0.43				0.44					
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person					0.70					
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal					0.66					
Displays a strong attachment for one particular member of the household					0.63					
Is active and energetic							0.87			
Is playful							0.90			

8.3.2 8M PCA extraction results

Table 8.5 8M PCA extraction for r-PWQ2 questions

Question	Extraction
Appears restless/agitated or paces when left, or about to be left	0.66
Whines when left, or about to be left	0.73
Barks when left, or about to be left	0.64
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	0.43
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk	0.82
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip	0.82
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	0.53
Is excessive, difficult to control and if it lunges is hard to hold back	0.53
Is self-controlled and calm	0.50
Is obviously disturbed by loud or unexpected sounds	0.72
Is spooked by odd or unexpected things or objects	0.78
Is anxious or uneasy in new situations	0.64
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)	0.56
Chases birds or squirrels (or would like to)	0.69
Chases cats (or would like to)	0.67
Pulls (including lunging) towards unfamiliar dogs	0.71
Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	0.51
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	0.72
Attention can be attracted easily but it loses interest soon	0.49
Will look at you when you talk to it directly in the home environment	0.53
Attention can be easily distracted	0.53
Needs obedience commands repeating to get a response	0.65
Seems like it doesn't listen even if it knows someone is speaking to it	0.64
Stays/waits when instructed to	0.50
Is uneasy with being physically handled/groomed	0.81
Attempts to move away when you start to groom it	0.80

Tends to follow you (or other member of household) about the house from room to room	0.62
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	0.49
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person	0.63
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal	0.63
Displays a strong attachment for one particular member of the household	0.50
Is active and energetic	0.85
Is playful	0.86

Table 8.6 Correlation matrix for 8M PCA using Varimax rotation. Component loadings above 0.3 (whether positive or negative) are shown and those below 0.4 highlighted in red.

Question	Component									
	1	2	3	4	5	6	7	8	9	10
Appears restless/agitated or paces when left, or about to be left			0.78							
Whines when left, or about to be left			0.84							
Barks when left, or about to be left			0.78							
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone			0.51			0.35				
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk					0.87					
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip					0.88					
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	0.43				0.36					
Is excessive, difficult to control and if it lunges is hard to hold back	0.42			0.33						
Is self-controlled and calm								0.66		
Is obviously disturbed by loud or unexpected sounds		0.83								
Is spooked by odd or unexpected things or objects		0.87								
Is anxious or uneasy in new situations		0.75								
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)		0.73								
Chases birds or squirrels (or would like to)	0.75									0.32
Chases cats (or would like to)	0.71									0.34
Pulls (including lunging) towards unfamiliar dogs	0.78									
Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	0.57									
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	0.77									
Attention can be attracted easily but it loses interest soon				0.60						
Will look at you when you talk to it directly in the home environment								0.69		

Attention can be easily distracted	0.34			0.59						
Needs obedience commands repeating to get a response				0.74						
Seems like it doesn't listen even if it knows someone is speaking to it				0.72						
Stays/waits when instructed to								0.69		
Is uneasy with being physically handled/groomed									0.88	
Attempts to move away when you start to groom it									0.86	
Tends to follow you (or other member of household) about the house from room to room						0.76				
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down						0.56				
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person										0.71
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal										0.73
Displays a strong attachment for one particular member of the household						0.66				
Is active and energetic							0.91			
Is playful							0.92			

8.3.3 12M PCA extraction results

Table 8.7 12M PCA extraction for r-PWQ2 questions.

Question	Extraction
Appears restless/agitated or paces when left, or about to be left	0.65
Whines when left, or about to be left	0.73
Barks when left, or about to be left	0.62
Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left	0.44
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	0.48
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk	0.85
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip	0.83
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home	0.51
Is excessive, difficult to control and if it lunges is hard to hold back	0.60
Is hyperactive, restless, has trouble settling down	0.50
Barks persistently when alarmed or excited	0.49
Is self-controlled and calm	0.45
Is obviously disturbed by loud or unexpected sounds	0.68
Is spooked by odd or unexpected things or objects	0.75
Is anxious or uneasy in new situations	0.66
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)	0.65
Chases birds or squirrels (or would like to)	0.66
Chases cats (or would like to)	0.67
Pulls (including lunging) towards unfamiliar dogs	0.67
Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public	0.50
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs	0.71
Attention can be attracted easily but it loses interest soon	0.45

Will look at you when you talk to it directly in the home environment	0.45
Attention can be easily distracted	0.48
Needs obedience commands repeating to get a response	0.61
Seems like it doesn't listen even if it knows someone is speaking to it	0.61
Stays/waits when instructed to	0.46
Is uneasy with being physically handled/groomed	0.81
Attempts to move away when you start to groom it	0.81
Tends to follow you (or other member of household) about the house from room to room	0.60
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down	0.53
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person	0.59
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal	0.53
Displays a strong attachment for one particular member of the household	0.48
Is active and energetic	0.82
Is playful	0.82

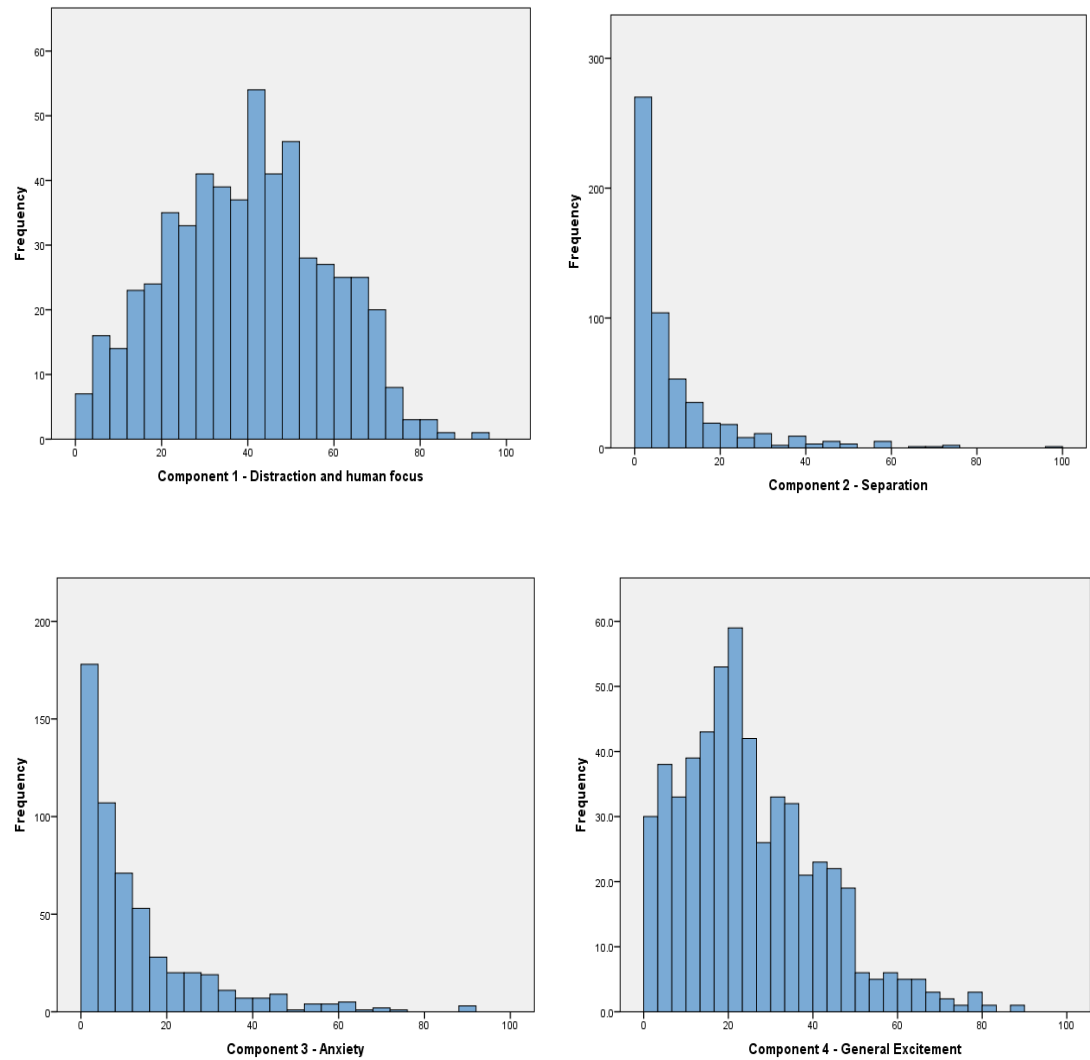
Table 8.8 Correlation matrix for 12M PCA using Varimax rotation. Component loadings above 0.3 (whether positive or negative) are shown and those below 0.4 shown in red. Questions that appear in two components are highlighted in bold where both loadings are above 0.4.

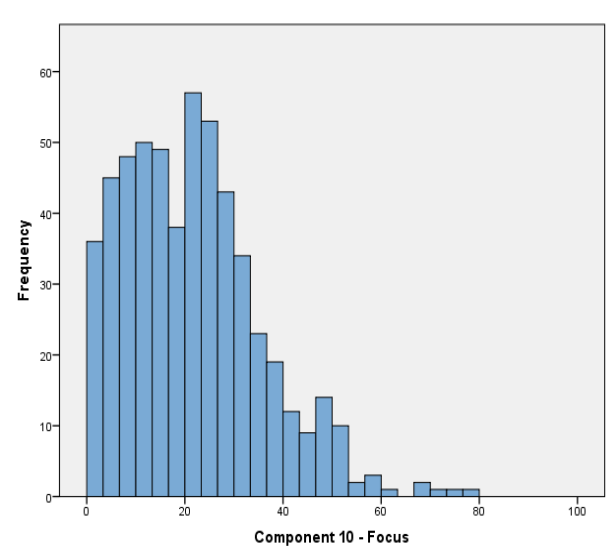
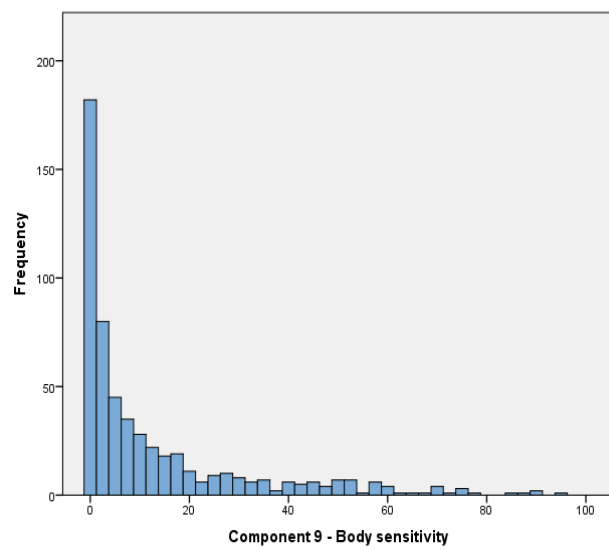
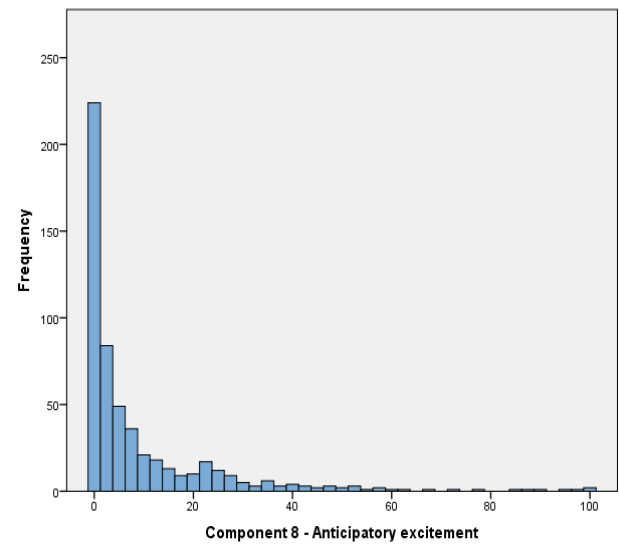
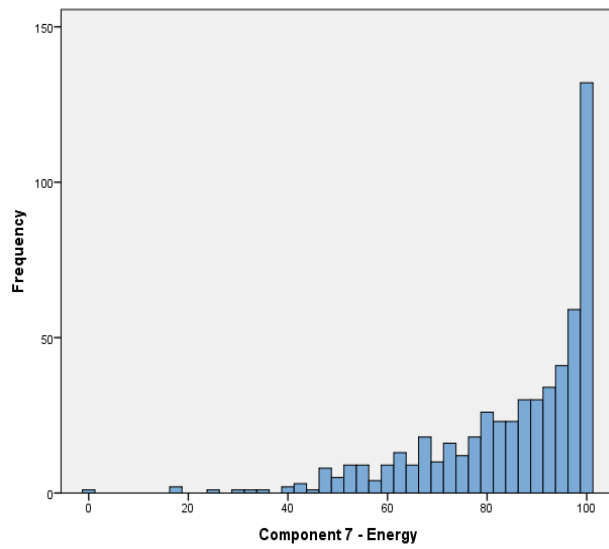
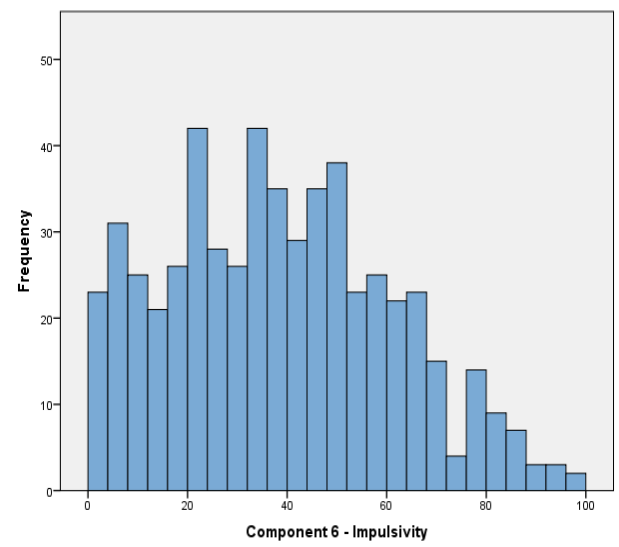
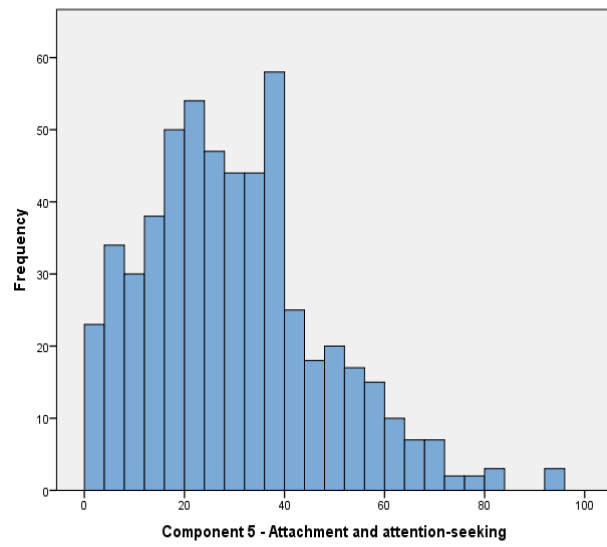
Question	Component									
	1	2	3	4	5	6	7	8	9	10
Appears restless/agitated or paces when left, or about to be left	0.77									
Whines when left, or about to be left	0.83									
Barks when left, or about to be left	0.77									
Chews/scratches at doors, floor, windows, curtains etc. when left, or about to be left	0.61									
Appears agitated (whines, barks, howls, scratches at door etc.) when separated from you (or a member of the household) but not alone	0.60									
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken for a walk							0.88			
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) just before being taken on a car trip							0.87			
Exhibits a high degree of excitement (jumps up, barks, coughs etc.) when visitors arrive at your home			0.37			0.50				
Is excessive, difficult to control and if it lunges is hard to hold back			0.38			0.56				
Is hyperactive, restless, has trouble settling down						0.57				
Barks persistently when alarmed or excited						0.63				
Is self-controlled and calm						-0.60				
Is obviously disturbed by loud or unexpected sounds		0.79								
Is spooked by odd or unexpected things or objects		0.84								
Is anxious or uneasy in new situations		0.75								
Backs away from or is reluctant to pass objects on the street (such as collecting boxes, bin bags or children's ride-on toys)		0.78								
Chases birds or squirrels (or would like to)			0.76							
Chases cats (or would like to)			0.76							
Pulls (including lunging) towards unfamiliar dogs			0.68	0.34						

Shows interest (attempts to greet, sniffs, wags tail) when passing children or members of the public			0.40	0.47						
Shows interest (attempts to greet, sniffs, wags tail) when encounters other dogs			0.68	0.42						
Attention can be attracted easily but it loses interest soon				0.59						
Will look at you when you talk to it directly in the home environment										0.61
Attention can be easily distracted				0.62						
Needs obedience commands repeating to get a response				0.69						
Seems like it doesn't listen even if it knows someone is speaking to it				0.66						
Stays/waits when instructed to										0.63
Is uneasy with being physically handled/groomed									0.88	
Attempts to move away when you start to groom it									0.87	
Tends to follow you (or other member of household) about the house from room to room					0.72					
Tends to nudge, nuzzle, or paw you (or others) for attention when you are sitting down				0.34	0.58					
Becomes agitated (whines, jumps up, tries to intervene) when you (or others) show affection for another person					0.51					-0.44
Becomes agitated (whines, jumps up, tries to intervene) when you show affection for another dog or animal					0.48					-0.34
Displays a strong attachment for one particular member of the household					0.65					
Is active and energetic									0.89	
Is playful									0.89	

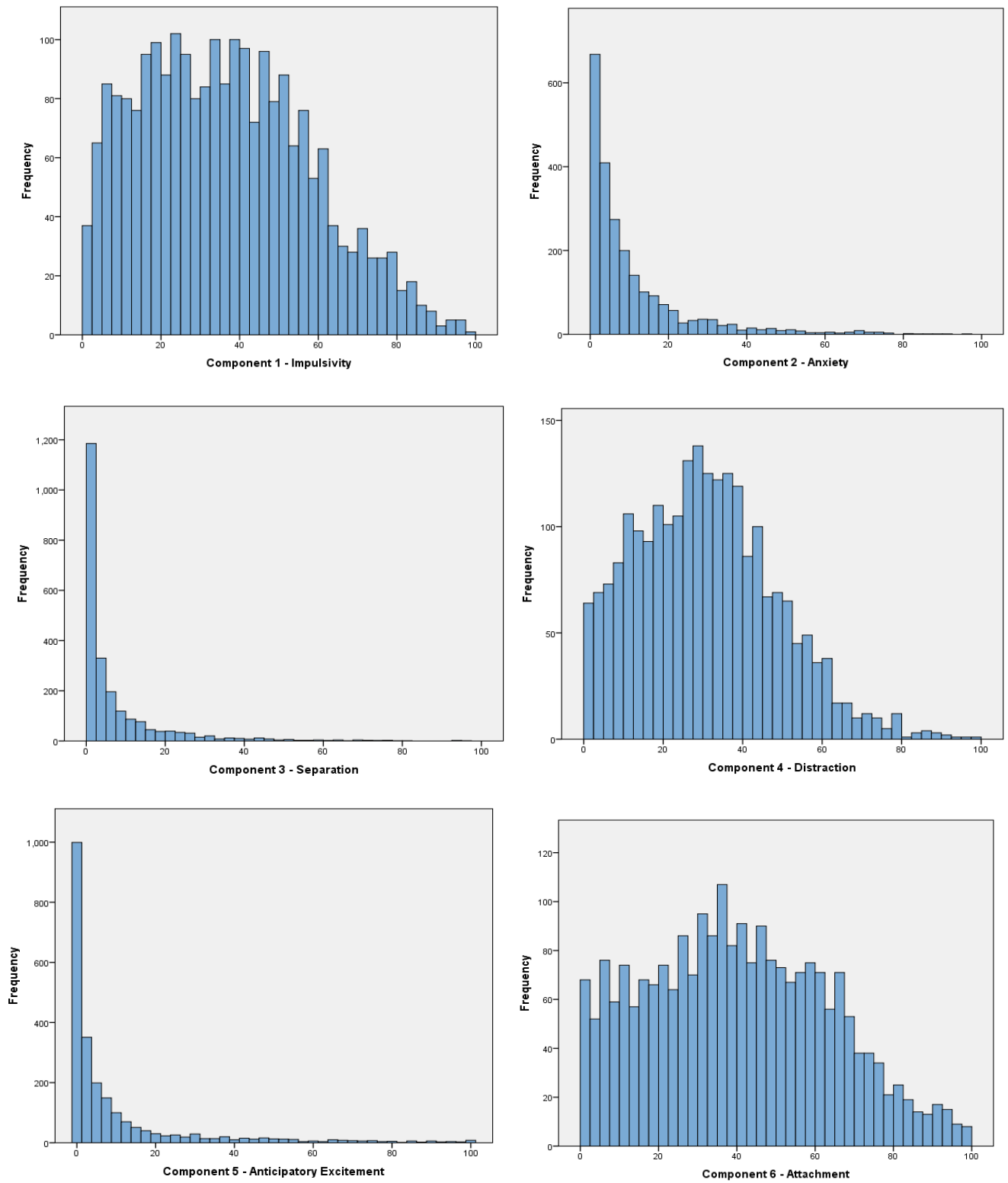
8.4 r-PWQ2 component distribution

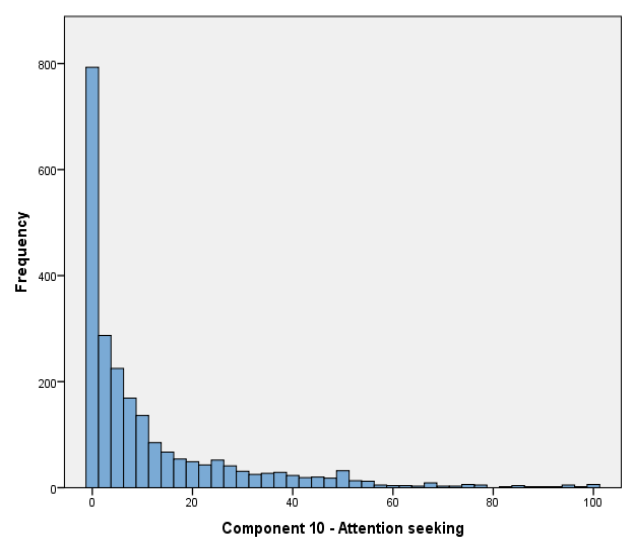
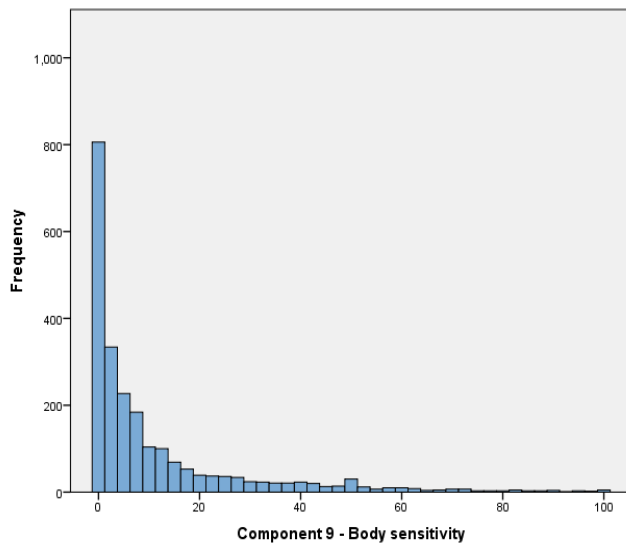
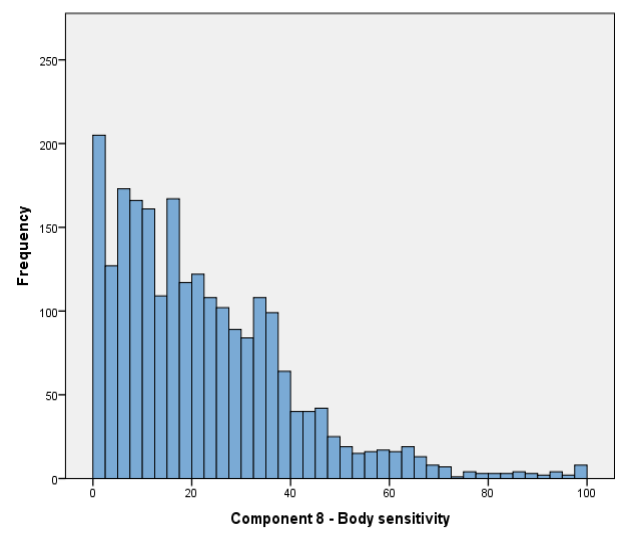
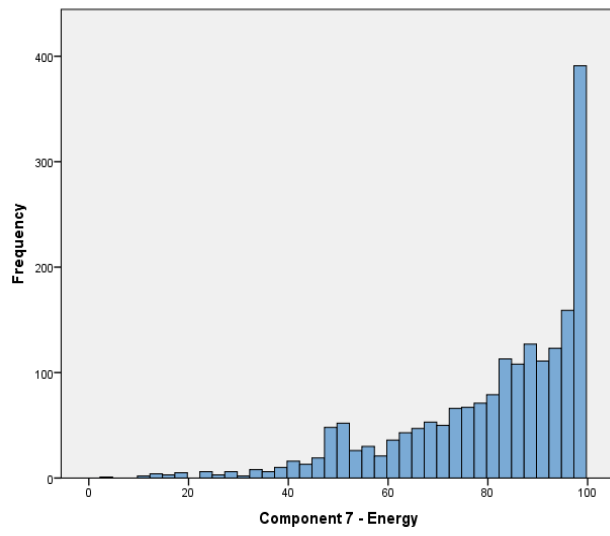
8.4.1 5M component r-PWQ2 distributions



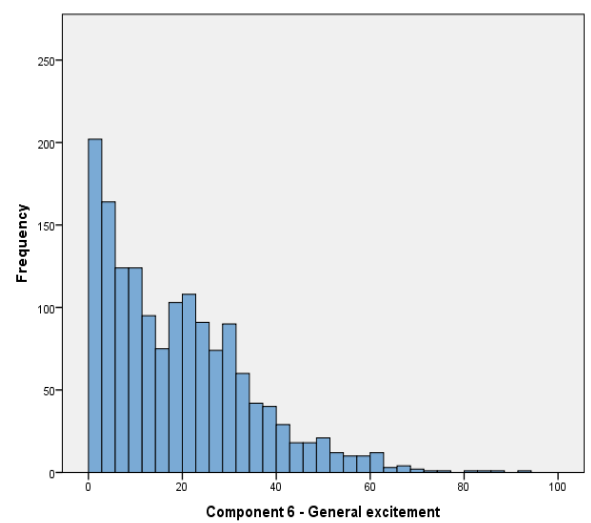
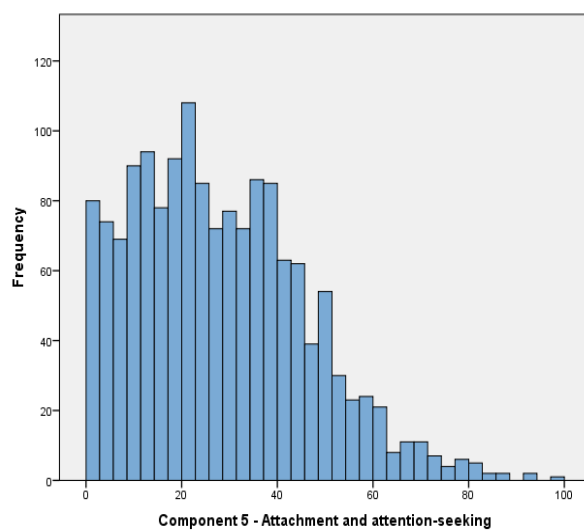
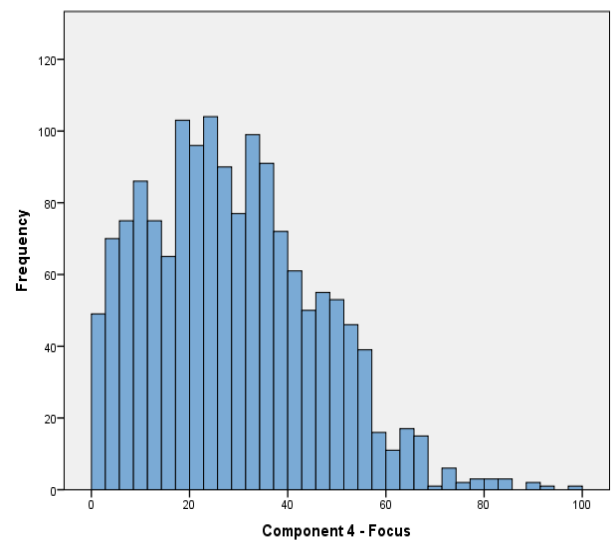
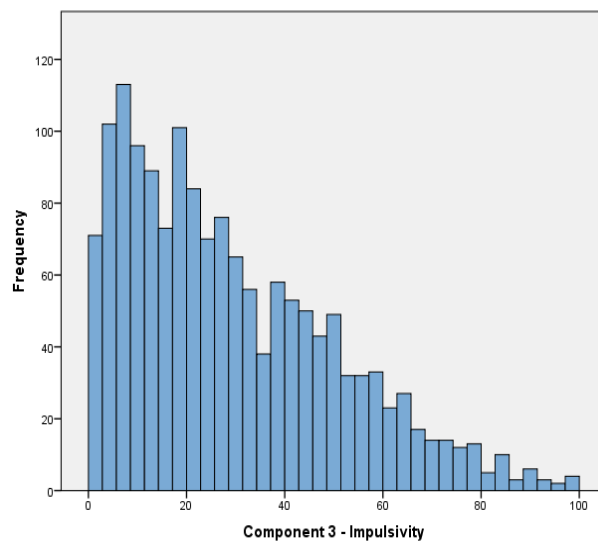
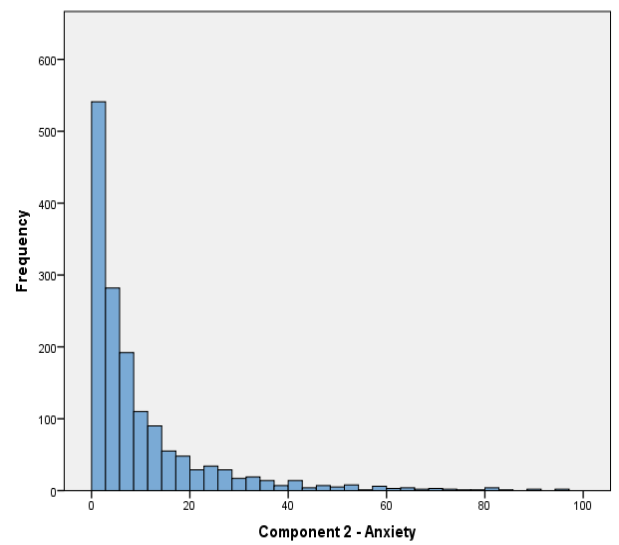
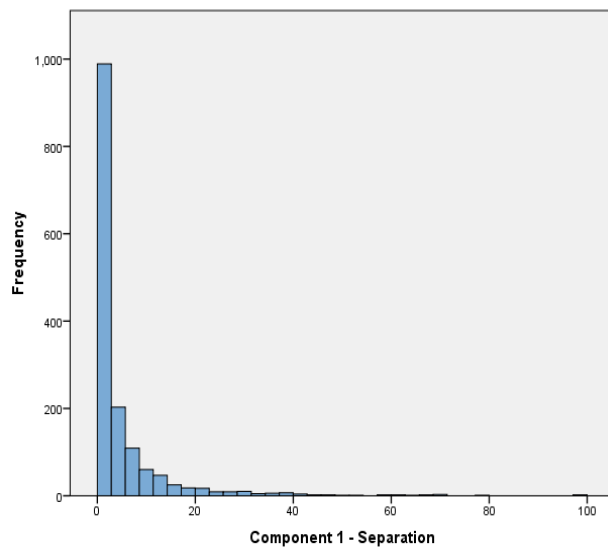


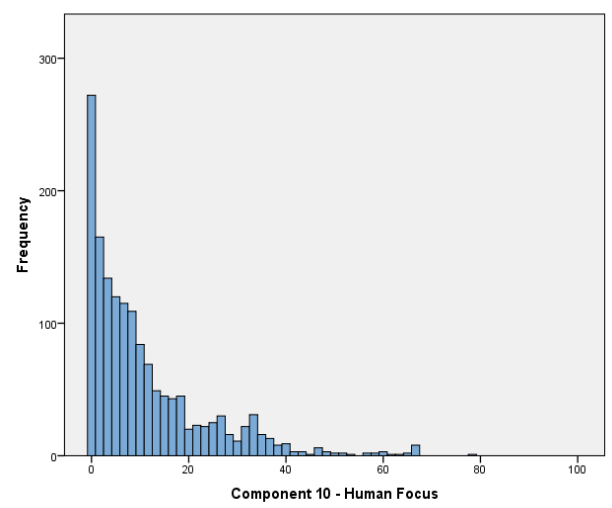
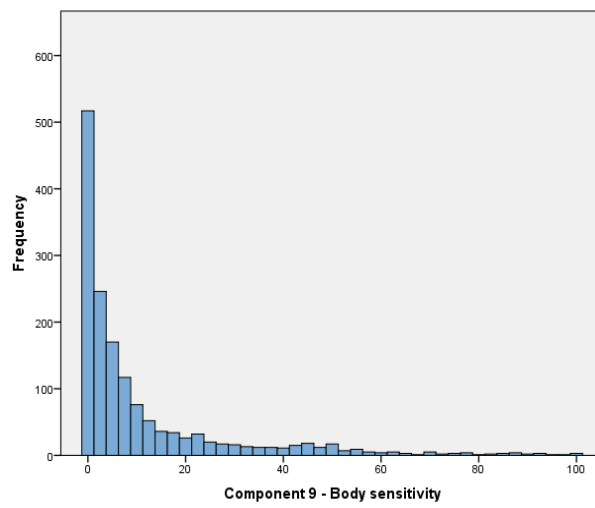
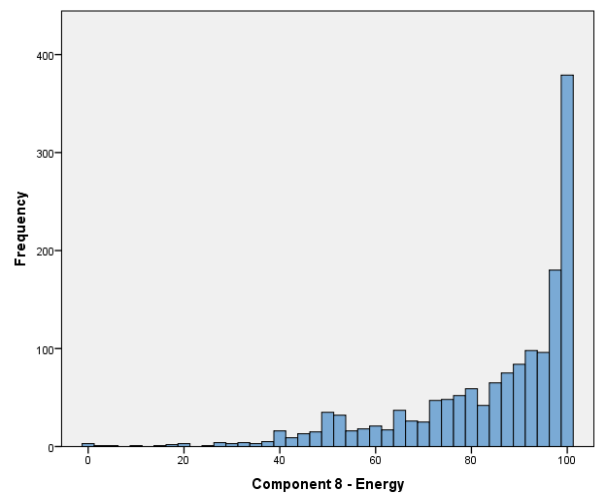
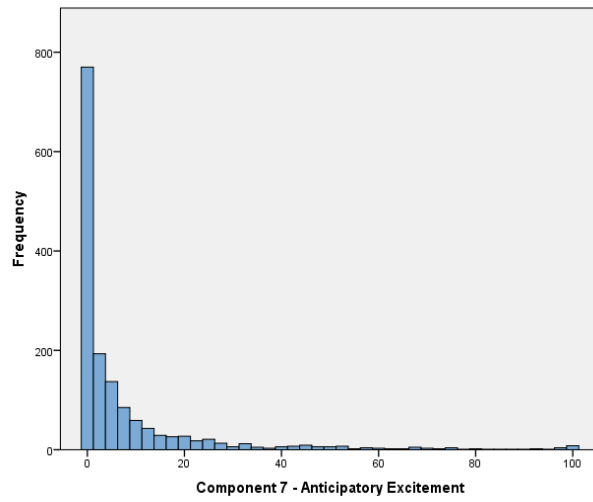
8.4.2 8M component r-PWQ2 distributions





8.4.3 12M component r-PWQ2 distributions





8.5 r-PWQ2 factor averages

8.5.1 5M factor averages

Table 8.9 Influence of sex and breed on 5M r-PWQ2 scores. Mean ($\pm$ S.D) for components 1-10. The lowest score for the factor is highlighted.

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Sex	Female	41.2 ($\pm$ 17.8)	8.1 ($\pm$ 11.7)	13.1 ($\pm$ 15.7)	26.1 ($\pm$ 16.9)	30.4 ($\pm$ 17.0)	39.8 ($\pm$ 22.2)	85.9 ($\pm$ 15.8)	9.7 ($\pm$ 16.9)	12.3 ($\pm$ 17.5)	21.6 ($\pm$ 13.3)
	Male	38.4 ($\pm$ 18.5)	9.5 ($\pm$ 14.8)	12.3 ($\pm$ 14.4)	24.1 ($\pm$ 15.8)	39.1 ($\pm$ 18.3)	37.1 ($\pm$ 22.6)	83.9 ($\pm$ 17.3)	9.0 ($\pm$ 15.1)	13.2 ($\pm$ 19.6)	21.4 ($\pm$ 14.7)
Breed	GR	40.4 ($\pm$ 17.9)	8.8 ($\pm$ 14.1)	16.6 ($\pm$ 18.6)	24.2 ($\pm$ 15.5)	27.7 ($\pm$ 16.1)	36.5 ($\pm$ 22.9)	84.7 ($\pm$ 17.7)	10.4 ($\pm$ 16.9)	13.8 ($\pm$ 21.0)	22.6 ($\pm$ 13.6)
	GR x GR* / GR x Lab*	39.7 ($\pm$ 19.9)	11.0 ($\pm$ 19.2)	17.7 ($\pm$ 17.0)	24.1 ($\pm$ 17.1)	29.5 ($\pm$ 16.6)	40.2 ($\pm$ 24.7)	81.8 ($\pm$ 18.0)	8.5 ($\pm$ 16.1)	10.1 ($\pm$ 15.1)	22.7 ($\pm$ 18.1)
	GR x GSD	28.9 ($\pm$ 13.6)	10.9 ($\pm$ 15.4)	18.9 ($\pm$ 11.9)	21.3 ($\pm$ 16.3)	22.9 ($\pm$ 17.4)	32.2 ($\pm$ 22.2)	89.6 ($\pm$ 13.2)	8.7 ($\pm$ 12.7)	13.0 ($\pm$ 21.8)	18.2 ($\pm$ 7.8)
	GR x Lab	39.7 ($\pm$ 18.1)	5.9 ($\pm$ 7.8)	9.5 ($\pm$ 12.1)	23.9 ($\pm$ 15.6)	27.8 ($\pm$ 16.2)	37.2 ($\pm$ 22.0)	82.6 ($\pm$ 19.1)	7.1 ($\pm$ 15.5)	11.9 ($\pm$ 17.2)	21.0 ($\pm$ 13.4)
	GSD	37.1 ($\pm$ 17.1)	14.5 ($\pm$ 17.1)	14.7 ($\pm$ 17.1)	31.9 ($\pm$ 19.5)	34.4 ($\pm$ 22.2)	38.9 ($\pm$ 21.1)	84.3 ($\pm$ 15.3)	8.1 ($\pm$ 13.6)	21.4 ($\pm$ 25.0)	23.8 ($\pm$ 13.7)
	Lab	42.5 ($\pm$ 16.9)	8.1 ($\pm$ 11.7)	9.3 ($\pm$ 11.5)	25.3 ($\pm$ 15.5)	31.0 ($\pm$ 17.7)	41.6 ($\pm$ 20.7)	87.2 ($\pm$ 14.1)	11.7 ($\pm$ 18.4)	11.3 ($\pm$ 16.6)	19.3 ($\pm$ 12.6)
	Lab x GR	37.8 ($\pm$ 22.3)	8.2 ($\pm$ 11.4)	13.2 ($\pm$ 17.3)	26.7 ($\pm$ 18.6)	32.4 ($\pm$ 20.2)	36.0 ($\pm$ 27.0)	87.5 ($\pm$ 15.7)	8.3 ($\pm$ 10.4)	11.8 ($\pm$ 18.6)	26.0 ($\pm$ 16.4)
	Lab x GR* / Lab x Lab*	30.8 ($\pm$ 15.7)	7.9 ($\pm$ 9.4)	14.9 ($\pm$ 11.3)	18.7 ($\pm$ 12.7)	30.1 ($\pm$ 18.1)	29.7 ($\pm$ 21.3)	84.9 ($\pm$ 14.4)	8.9 ($\pm$ 9.2)	12.4 ($\pm$ 14.2)	17.9 ($\pm$ 10.8)
	SP x Lab / Lab x SP	43.9 ($\pm$ 15.5)	19.1 ($\pm$ 26.8)	11.3 ($\pm$ 16.3)	35.3 ($\pm$ 17.0)	23.9 ($\pm$ 18.6)	49.1 ($\pm$ 13.3)	91.5 ($\pm$ 8.5)	4.4 ($\pm$ 9.6)	19.5 ($\pm$ 21.8)	32.1 ($\pm$ 11.9)

Table 8.10 Influence of additional factors on 5M r-PWQ2 scores. Mean ($\pm$ standard deviation) for factor categories for components 1-10 at 5M. The lowest score for the factor is highlighted

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
PR experience	< 5	43.6 ($\pm$ 18.2)	9.2 ($\pm$ 13.7)	12.3 ($\pm$ 14.9)	26.1 ($\pm$ 16.3)	30.5 ($\pm$ 17.1)	42.0 ($\pm$ 22.0)	85.0 ($\pm$ 16.4)	9.5 ($\pm$ 17.1)	12.2 ($\pm$ 17.5)	21.3 ($\pm$ 14.3)
	5 to 9	34.3 ($\pm$ 17.7)	8.1 ($\pm$ 12.7)	11.4 ($\pm$ 12.8)	23.1 ($\pm$ 17.2)	30.8 ($\pm$ 18.6)	32.6 ($\pm$ 23.0)	85.4 ($\pm$ 17.6)	9.7 ($\pm$ 15.4)	13.0 ($\pm$ 19.2)	21.6 ($\pm$ 14.0)
	10 to 19	36.9 ($\pm$ 16.1)	9.3 ($\pm$ 14.7)	17.8 ($\pm$ 19.3)	26.2 ($\pm$ 15.7)	30.0 ($\pm$ 18.7)	36.4 ($\pm$ 22.5)	85.3 ($\pm$ 14.1)	9.8 ($\pm$ 14.9)	13.6 ($\pm$ 22.2)	22.6 ($\pm$ 11.6)
	20 +	30.1 ($\pm$ 12.8)	6.5 ($\pm$ 8.7)	11.8 ($\pm$ 12.8)	20.4 ($\pm$ 14.0)	17.7 ($\pm$ 13.6)	30.6 ($\pm$ 17.4)	81.8 ($\pm$ 19.1)	5.7 ($\pm$ 7.5)	15.8 ($\pm$ 18.8)	21.5 ($\pm$ 15.6)
Adults	1	37.8 ($\pm$ 18.8)	7.2 ($\pm$ 10.1)	11.1 ($\pm$ 13.8)	27.7 ($\pm$ 19.6)	28.9 ($\pm$ 18.0)	34.7 ($\pm$ 21.5)	85.5 ($\pm$ 16.1)	8.1 ($\pm$ 15.3)	15.2 ($\pm$ 20.8)	22.5 ($\pm$ 14.3)
	2	39.9 ($\pm$ 18.4)	8.7 ($\pm$ 13.7)	12.8 ($\pm$ 15.5)	24.5 ($\pm$ 15.8)	29.5 ($\pm$ 18.0)	38.9 ($\pm$ 22.6)	85.7 ($\pm$ 16.2)	9.2 ($\pm$ 15.8)	12.3 ($\pm$ 18.3)	21.0 ($\pm$ 14.0)
	3	40.0 ($\pm$ 17.8)	11.0 ($\pm$ 14.5)	12.3 ($\pm$ 13.0)	24.4 ($\pm$ 15.8)	32.1 ($\pm$ 16.5)	39.5 ($\pm$ 23.1)	83.3 ($\pm$ 16.4)	11.7 ($\pm$ 18.5)	12.4 ($\pm$ 17.3)	21.4 ($\pm$ 13.9)
	4	46.4 ($\pm$ 15.1)	7.5 ($\pm$ 10.4)	17.4 ($\pm$ 17.4)	25.0 ($\pm$ 15.0)	33.5 ($\pm$ 15.6)	40.3 ($\pm$ 20.2)	82.9 ($\pm$ 18.1)	10.8 ($\pm$ 16.6)	12.5 ($\pm$ 20.8)	22.5 ($\pm$ 13.1)
	5+	44.6 ($\pm$ 14.5)	3.5 ($\pm$ 1.8)	11.0 ($\pm$ 10.5)	31.6 ($\pm$ 12.8)	25.9 ($\pm$ 16.9)	48.5 ($\pm$ 25.3)	74.6 ($\pm$ 29.1)	7.5 ($\pm$ 12.2)	17.4 ($\pm$ 25.7)	28.8 ($\pm$ 11.8)
Children	Yes	45.2 ($\pm$ 18.0)	9.5 ($\pm$ 11.1)	14.2 ($\pm$ 16.2)	28.1 ($\pm$ 17.6)	32.2 ($\pm$ 17.5)	45.0 ($\pm$ 22.1)	83.4 ($\pm$ 16.2)	11.9 ($\pm$ 19.2)	14.2 ($\pm$ 19.3)	24.6 ($\pm$ 14.3)
	No	38.8 ($\pm$ 17.3)	7.3 ($\pm$ 12.0)	11.5 ($\pm$ 13.3)	24.6 ($\pm$ 16.7)	28.8 ($\pm$ 16.9)	37.6 ($\pm$ 22.3)	85.2 ($\pm$ 16.6)	8.1 ($\pm$ 15.0)	12.4 ($\pm$ 18.1)	19.9 ($\pm$ 13.3)
Dogs	Yes	36.1 ($\pm$ 17.0)	10.1 ($\pm$ 15.9)	12.9 ($\pm$ 16.8)	22.4 ($\pm$ 14.5)	28.5 ($\pm$ 18.0)	35.2 ($\pm$ 21.8)	86.7 ($\pm$ 15.0)	10.0 ($\pm$ 16.8)	9.3 ($\pm$ 14.1)	21.3 ($\pm$ 13.8)
	No	41.8 ($\pm$ 18.1)	7.9 ($\pm$ 11.4)	12.8 ($\pm$ 14.0)	26.5 ($\pm$ 16.7)	30.2 ($\pm$ 17.0)	40.0 ($\pm$ 22.3)	82.8 ($\pm$ 18.0)	8.9 ($\pm$ 15.9)	14.9 ($\pm$ 20.5)	21.1 ($\pm$ 13.5)
Animals	Yes	37.3 ($\pm$ 17.1)	9.0 ($\pm$ 14.4)	11.8 ($\pm$ 15.2)	22.9 ($\pm$ 15.2)	28.6 ($\pm$ 17.8)	36.9 ($\pm$ 22.3)	85.9 ($\pm$ 16.1)	10.1 ($\pm$ 17.3)	10.5 ($\pm$ 16.2)	21.6 ($\pm$ 14/3)

	No	42.2 (± 18.5)	8.3 (± 11.8)	13.4 (± 14.6)	27.1 (± 17.2)	30.9 (± 17.3)	39.8 (± 22.4)	83.6 (± 17.2)	8.6 (± 14.8)	15.1 (± 20.6)	21.3 (± 13.3)
Location	Rural	39.5 (± 17.8)	8.1 (± 13.5)	13.0 (± 13.6)	24.2 (± 15.4)	29.9 (± 17.8)	37.6 (± 21.5)	84.1 (± 17.6)	11.0 (± 18.4)	10.0 (± 15.0)	21.4 (± 14.5)
	Suburban	39.4 (± 18.3)	9.2 (± 13.8)	12.1 (± 15.0)	25.0 (± 17.1)	28.7 (± 17.4)	37.9 (± 22.3)	86.1 (± 15.8)	8.0 (± 13.9)	13.1 (± 19.5)	21.4 (± 13.8)
	Urban	41.4 (± 18.4)	8.1 (± 10.1)	14.2 (± 17.2)	27.0 (± 15.9)	33.6 (± 18.1)	41.1 (± 24.0)	82.6 (± 16.4)	10.9 (± 17.4)	17.1 (± 20.8)	22.6 (± 13.6)
Puppy Class	Yes	40.3 (± 18.0)	8.8 (± 12.8)	12.4 (± 14.6)	25.8 (±16.5)	29.9 (± 17.6)	39.2 (± 22.2)	84.5 (± 16.4)	9.6 (± 15.9)	13.7 (± 19.6)	22.0 (± 14.1)
	No	37.1 (± 18.5)	8.4 (± 15.1)	14.3 (± 17.2)	21.6 (±15.3)	29.5 (± 18.3)	34.6 (± 22.7)	86.6 (± 17.4)	8.5 (± 16.9)	8.1 (± 12.2)	19.7 (± 13.5)
Work	Yes	38.1 (± 21.2)	8.4 (± 16.5)	8.4 (± 11.6)	22.4 (± 15.2)	30.5 (± 17.3)	35.0 (± 22.9)	85.9 (± 16.0)	4.3 (± 7.2)	8.2 (± 13.0)	21.4 (± 18.3)
	No	40.0 (±17.7)	8.7 (± 12.7)	13.3 (± 15.3)	25.4 (±16.5)	29.7 (± 17.8)	38.8 (± 22.3)	84.8 (± 16.7)	10.1 (± 16.8)	13.3 (± 19.0)	21.7 (± 13.4)

8.5.2 8M factor averages

Table 8.11 Influence of sex and breed on 8M r-PWQ2 scores. Mean ($\pm$ S.D) for components 1-10. The lowest score for the factor is highlighted

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Sex	Female	37.3 ($\pm$ 21.8)	11.2 ($\pm$ 14.6)	6.4 ($\pm$ 11.2)	31.2 ($\pm$ 18.2)	10.0 ($\pm$ 18.6)	39.4 ($\pm$ 23.1)	85.4 ($\pm$ 16.9)	22.5 ($\pm$ 17.5)	11.7 ($\pm$ 18.3)	11.7 ($\pm$ 16.9)
	Male	35.4 ($\pm$ 20.8)	9.9 ($\pm$ 13.1)	7.4 ($\pm$ 12.6)	30.4 ($\pm$ 17.3)	9.6 ($\pm$ 17.5)	41.5 ($\pm$ 24.0)	82.7 ($\pm$ 19.1)	22.2 ($\pm$ 18.4)	10.7 ($\pm$ 16.6)	11.8 ($\pm$ 17.9)
Breed	CCR + CCR x Lab	38.3 ($\pm$ 22.3)	7.4 ($\pm$ 4.6)	7.6 ($\pm$ 6.9)	41.3 ($\pm$ 20.2)	2.2 ($\pm$ 3.1)	41.9 ($\pm$ 16.0)	73.0 ($\pm$ 25.5)	21.6 ($\pm$ 16.0)	27.0 ($\pm$ 25.3)	22.9 ($\pm$ 20.0)
	GR	33.0 ($\pm$ 21.0)	14.6 ($\pm$ 18.0)	6.0 ($\pm$ 10.8)	30.9 ($\pm$ 18.3)	9.0 ($\pm$ 17.3)	35.2 ($\pm$ 23.8)	81.1 ($\pm$ 18.8)	23.9 ($\pm$ 19.3)	10.4 ($\pm$ 17.7)	11.6 ($\pm$ 18.4)
	GR x GR* + GR x Lab*	38.0 ($\pm$ 21.9)	12.5 ($\pm$ 13.3)	7.4 ($\pm$ 13.1)	32.0 ($\pm$ 17.0)	9.0 ($\pm$ 17.1)	40.1 ($\pm$ 22.7)	81.1 ($\pm$ 18.7)	24.0 ($\pm$ 19.1)	9.1 ($\pm$ 15.8)	11.6 ($\pm$ 17.9)
	GR x GSD	37.5 ($\pm$ 22.6)	14.8 ($\pm$ 17.5)	7.4 ($\pm$ 12.6)	30.1 ($\pm$ 18.0)	11.4 ($\pm$ 18.9)	44.2 ($\pm$ 22.1)	87.0 ($\pm$ 14.1)	21.2 ($\pm$ 16.1)	8.0 ($\pm$ 12.0)	18.7 ($\pm$ 24.1)
	GR x Lab	36.8 ($\pm$ 21.3)	11.3 ($\pm$ 15.0)	6.8 ($\pm$ 11.3)	30.8 ($\pm$ 17.9)	9.0 ($\pm$ 17.7)	40.8 ($\pm$ 23.2)	84.7 ($\pm$ 18.1)	20.9 ($\pm$ 16.4)	11.5 ($\pm$ 18.0)	10.7 ($\pm$ 16.0)
	GSD	36.3 ($\pm$ 22.7)	10.6 ($\pm$ 13.3)	8.7 ($\pm$ 11.6)	33.9 ($\pm$ 19.1)	10.1 ($\pm$ 16.6)	38.6 ($\pm$ 22.8)	86.6 ($\pm$ 17.8)	26.7 ($\pm$ 20.0)	9.5 ($\pm$ 14.4)	12.8 ($\pm$ 16.4)
	Lab	37.7 ($\pm$ 20.5)	7.8 ($\pm$ 10.6)	6.8 ($\pm$ 12.1)	30.6 ($\pm$ 17.6)	11.8 ($\pm$ 20.3)	41.7 ($\pm$ 23.9)	84.4 ($\pm$ 17.8)	22.9 ($\pm$ 18.3)	12.2 ($\pm$ 18.4)	11.6 ($\pm$ 17.1)
	Lab x GR	34.5 ($\pm$ 21.3)	9.6 ($\pm$ 13.3)	6.3 ($\pm$ 18.9)	29.0 ($\pm$ 16.6)	8.0 ($\pm$ 14.9)	39.5 ($\pm$ 23.8)	84.5 ($\pm$ 18.2)	20.0 ($\pm$ 17.1)	11.5 ($\pm$ 17.5)	11.2 ($\pm$ 16.6)
	Lab x GR* + Lab x Lab*	35.2 ($\pm$ 20.8)	9.5 ($\pm$ 11.1)	7.3 ($\pm$ 12.6)	29.9 ($\pm$ 17.6)	10.2 ($\pm$ 18.1)	43.4 ($\pm$ 24.1)	84.4 ($\pm$ 17.8)	22.4 ($\pm$ 18.1)	11.7 ($\pm$ 17.6)	12.6 ($\pm$ 17.5)
	SP x Lab + Lab x SP	34.6 ($\pm$ 26.1)	7.3 ($\pm$ 9.5)	9.3 ($\pm$ 14.0)	29.5 ($\pm$ 20.8)	7.9 ($\pm$ 13.7)	40.0 ($\pm$ 27.0)	87.5 ($\pm$ 19.4)	19.9 ($\pm$ 17.6)	11.4 ($\pm$ 18.1)	9.5 ($\pm$ 17.9)

Table 8.12 Influence of additional factors on 8M r-PWQ2 scores. Mean ($\pm$ standard deviation) for factor categories for components 1-10 at 8M. The lowest score for the factor is highlighted

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
PR experience	< 5	39.6 ($\pm$ 20.9)	10.2 ($\pm$ 13.3)	7.4 ($\pm$ 12.3)	31.9 ($\pm$ 17.5)	9.4 ($\pm$ 17.7)	42.5 ($\pm$ 23.2)	83.7 ($\pm$ 18.1)	21.5 ($\pm$ 17.0)	11.7 ($\pm$ 18.0)	11.3 ($\pm$ 17.1)
	5 to 9	32.3 ($\pm$ 20.3)	11.2 ($\pm$ 14.2)	6.4 ($\pm$ 11.1)	29.0 ($\pm$ 17.3)	11.3 ($\pm$ 19.8)	38.7 ($\pm$ 23.4)	84.6 ($\pm$ 17.0)	23.1 ($\pm$ 19.4)	9.7 ($\pm$ 15.4)	12.0 ($\pm$ 17.5)
	10 to 19	31.1 ($\pm$ 21.2)	11.5 ($\pm$ 15.1)	6.1 ($\pm$ 11.1)	28.2 ($\pm$ 18.4)	9.7 ($\pm$ 16.8)	38.4 ($\pm$ 24.8)	84.5 ($\pm$ 19.5)	22.4 ($\pm$ 18.0)	10.5 ($\pm$ 16.4)	13.2 ($\pm$ 18.3)
	20 +	29.1 ($\pm$ 22.2)	8.6 ($\pm$ 13.3)	6.1 ($\pm$ 12.4)	31.8 ($\pm$ 19.1)	8.5 ($\pm$ 16.0)	30.2 ($\pm$ 21.9)	84.4 ($\pm$ 19.3)	28.5 ($\pm$ 21.0)	14.9 ($\pm$ 22.5)	12.8 ($\pm$ 17.5)
Adults	1	36.2 ($\pm$ 20.9)	13.0 ($\pm$ 16.0)	6.6 ($\pm$ 12.3)	32.3 ($\pm$ 18.3)	9.5 ($\pm$ 17.1)	43.1 ($\pm$ 21.6)	84.9 ($\pm$ 18.8)	22.5 ($\pm$ 17.2)	13.6 ($\pm$ 20.1)	14.1 ($\pm$ 19.0)
	2	35.9 ($\pm$ 21.1)	10.3 ($\pm$ 13.4)	7.2 ($\pm$ 12.6)	31.3 ($\pm$ 17.3)	8.5 ($\pm$ 17.0)	40.5 ($\pm$ 23.7)	85.7 ($\pm$ 16.6)	20.7 ($\pm$ 16.3)	11.1 ($\pm$ 17.4)	11.6 ($\pm$ 17.0)
	3	35.8 ($\pm$ 22.6)	10.9 ($\pm$ 14.0)	7.7 ($\pm$ 12.0)	29.5 ($\pm$ 17.8)	10.6 ($\pm$ 18.7)	38.3 ($\pm$ 22.5)	84.4 ($\pm$ 18.2)	19.8 ($\pm$ 16.0)	11.1 ($\pm$ 16.9)	12.7 ($\pm$ 19.8)
	4	37.0 ($\pm$ 22.0)	14.6 ($\pm$ 16.8)	9.0 ($\pm$ 12.1)	33.7 ($\pm$ 18.2)	8.1 ($\pm$ 14.9)	41.3 ($\pm$ 22.7)	82.4 ($\pm$ 17.4)	22.8 ($\pm$ 18.6)	12.2 ($\pm$ 17.4)	10.2 ($\pm$ 14.2)
	5+	34.5 ($\pm$ 24.1)	9.6 ($\pm$ 13.5)	5.0 ($\pm$ 6.3)	31.0 ($\pm$ 19.0)	6.9 ($\pm$ 17.2)	37.9 ($\pm$ 23.7)	82.2 ($\pm$ 21.2)	16.7 ($\pm$ 14.8)	11.3 ($\pm$ 19.0)	8.5 ($\pm$ 12.4)
Children	Yes	38.6 ($\pm$ 21.5)	12.1 ($\pm$ 15.4)	8.1 ($\pm$ 13.0)	32.0 ($\pm$ 17.5)	11.6 ($\pm$ 20.2)	44.1 ($\pm$ 22.9)	80.8 ($\pm$ 19.0)	20.6 ($\pm$ 15.3)	11.5 ($\pm$ 16.8)	14.3 ($\pm$ 20.0)
	No	35.6 ($\pm$ 20.9)	10.8 ($\pm$ 14.2)	6.7 ($\pm$ 12.3)	31.2 ($\pm$ 17.6)	8.4 ($\pm$ 16.6)	40.5 ($\pm$ 23.5)	86.3 ($\pm$ 16.7)	20.3 ($\pm$ 16.4)	11.5 ($\pm$ 18.1)	11.2 ($\pm$ 16.4)
Dogs	Yes	33.7 ($\pm$ 21.9)	11.5 ($\pm$ 14.9)	7.3 ($\pm$ 12.4)	30.4 ($\pm$ 18.0)	10.2 ($\pm$ 18.2)	37.2 ($\pm$ 23.6)	86.2 ($\pm$ 16.6)	21.6 ($\pm$ 17.5)	10.4 ($\pm$ 16.6)	15.2 ($\pm$ 19.0)
	No	37.3 ($\pm$ 20.9)	10.4 ($\pm$ 13.4)	7.0 ($\pm$ 12.1)	31.7 ($\pm$ 17.5)	8.2 ($\pm$ 16.7)	42.4 ($\pm$ 23.2)	84.7 ($\pm$ 17.5)	20.3 ($\pm$ 15.7)	12.3 ($\pm$ 18.6)	10.1 ($\pm$ 16.1)
Animals	Yes	34.6 ($\pm$ 21.9)	10.9 ($\pm$ 14.4)	7.3 ($\pm$ 12.3)	30.6 ($\pm$ 18.0)	9.8 ($\pm$ 17.9)	38.0 ($\pm$ 23.5)	86.1 ($\pm$ 16.8)	21.2 ($\pm$ 17.3)	10.4 ($\pm$ 16.8)	13.5 ($\pm$ 18.0)

	No	37.5 (± 20.6)	10.6 (± 13.5)	6.8 (± 12.2)	31.9 (± 17.4)	8.1 (± 16.5)	43.3 (± 23.2)	84.4 (± 17.6)	20.3 (± 15.4)	12.9 (± 19.0)	10.5 (± 16.8)
Location	Rural	34.2 (± 20.6)	11.1 (± 14.4)	6.6 (± 12.0)	31.4 (± 16.7)	8.7 (± 16.5)	39.9 (± 23.8)	85.4 (± 17.5)	20.5 (± 16.2)	11.1 (± 17.3)	11.8 (± 16.9)
	Suburban	36.1 (± 21.4)	10.2 (± 13.1)	7.1 (± 12.3)	30.6 (± 18.0)	8.4 (± 16.8)	39.4 (± 23.0)	85.2 (± 17.1)	20.5 (± 16.5)	11.1 (± 17.4)	11.0 (± 16.5)
	Urban	38.7 (± 22.6)	12.5 (± 15.9)	8.4 (± 12.9)	33.6 (± 18.0)	11.0 (± 19.6)	45.8 (± 24.5)	84.7 (± 18.0)	23.1 (± 17.4)	14.0 (± 20.1)	15.4 (± 21.0)
Puppy Class	Yes	37.4 (± 21.1)	10.1 (± 13.5)	6.9 (± 11.9)	31.0 (± 17.7)	9.5 (± 17.6)	41.6 (± 23.7)	83.2 (± 18.5)	22.5 (± 18.1)	11.4 (± 17.8)	10.8 (± 16.9)
	No	36.2 (± 20.5)	10.0 (± 14.1)	6.2 (± 10.4)	31.6 (± 19.3)	10.7 (± 21.2)	42.6 (± 23.0)	81.9 (± 19.8)	23.1 (± 20.5)	8.7 (± 14.7)	10.1 (± 15.3)
Work	Yes	35.9 (± 21.5)	11.2 (± 15.5)	6.9 (± 12.3)	33.2 (± 19.1)	8.1 (± 14.9)	41.7 (± 23.8)	82.6 (± 18.2)	23.4 (± 20.0)	11.5 (± 19.6)	13.0 (± 20.3)
	No	37.5 (± 21.0)	10.0 (± 13.2)	6.9 (± 11.8)	30.7 (± 17.6)	9.8 (± 18.3)	41.7 (± 23.6)	83.1 (± 18.8)	22.4 (± 18.1)	11.2 (± 17.3)	10.4 (± 16.1)
YOB	2016	39.1 (± 20.5)	9.5 (± 13.5)	5.5 (± 9.4)	27.9 (± 18.3)	15.7 (± 22.3)	39.5 (± 22.8)	78.0 (± 20.8)	29.3 (± 22.3)	9.2 (± 14.8)	10.4 (± 15.7)
	2017	34.8 (± 21.3)	10.7 (± 14.4)	7.1 (± 12.0)	30.6 (± 17.6)	8.7 (± 16.9)	40.9 (± 24.5)	85.0 (± 17.7)	22.2 (± 18.4)	11.0 (± 17.5)	12.0 (± 17.8)
	2018	36.0 (± 21.8)	11.0 (± 14.2)	7.0 (± 11.8)	31.3 (± 18.3)	8.7 (± 17.2)	39.8 (± 22.8)	84.8 (± 17.3)	19.4 (± 15.1)	11.6 (± 17.9)	11.5 (± 17.1)
	2019	37.8 (± 20.9)	10.4 (± 12.5)	7.9 (± 13.2)	32.1 (± 16.5)	9.7 (± 17.6)	41.3 (± 23.3)	84.9 (± 17.4)	22.1 (± 16.6)	12.4 (± 18.4)	12.6 (± 18.0)
Neutered	Yes	36.9 (± 21.1)	8.8 (± 11.2)	7.8 (± 12.1)	30.6 (± 17.8)	9.2 (± 16.7)	41.7 (± 23.5)	84.5 (± 18.1)	22.4 (± 18.8)	11.3 (± 17.1)	10.4 (± 16.1)
	No	36.4 (± 21.5)	11.2 (± 14.6)	6.9 (± 12.1)	30.9 (± 17.7)	10.0 (± 18.4)	40.5 (± 23.7)	83.9 (± 18.2)	22.4 (± 17.7)	11.2 (± 17.5)	12.2 (± 17.9)

8.5.3 12M factor averages

Table 8.13 Influence of sex and breed on 12M r-PWQ2 scores. Mean ($\pm$ S.D) for components 1-10. The lowest score for the factor is highlighted

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Sex	Female	4.5 (± 9.5)	10.2 (± 13.2)	29.5 (± 22.1)	28.7 (± 17.1)	27.5 (± 18.1)	18.7 (± 15.4)	8.4 (± 16.3)	83.9 (± 18.9)	11.1 (± 17.2)	10.9 (± 12.8)
	Male	5.0 (± 9.7)	9.6 (± 14.2)	28.9 (± 20.7)	29.0 (± 17.2)	28.7 (± 17.9)	19.3 (± 15.3)	7.6 (± 16.2)	84.1 (± 17.7)	11.0 (± 17.8)	11.0 (± 11.9)
Breed	CCR + CCR x Lab	5.6 (± 3.6)	17.7 (± 10.2)	52.5 (± 11.0)	41.3 (± 9.9)	36.9 (± 10.2)	32.9 (± 8.8)	6.7 (± 8.7)	82.5 (± 17.8)	17.1 (± 18.1)	12.6 (± 8.8)
	GR	3.8 (± 6.7)	15.9 (± 19.6)	29.7 (± 19.8)	32.0 (± 16.8)	27.8 (± 19.2)	17.4 (± 13.0)	9.0 (± 15.6)	82.0 (± 19.5)	9.1 (± 16.0)	12.2 (± 12.4)
	GR x GR* + GR x Lab*	4.1 (± 6.4)	12.4 (± 14.7)	29.2 (± 22.7)	30.5 (± 18.3)	28.5 (± 18.0)	17.9 (± 14.0)	6.6 (± 13.2)	78.7 (± 21.1)	11.3 (± 16.7)	13.1 (± 14.2)
	GR x GSD	6.5 (± 13.5)	10.8 (± 14.9)	30.9 (± 21.0)	26.6 (± 13.9)	27.9 (± 18.7)	18.4 (± 17.7)	7.0 (± 13.2)	84.0 (± 18.2)	6.1 (± 10.8)	9.1 (± 9.6)
	GR x Lab	4.9 (± 10.2)	10.0 (± 13.9)	28.9 (± 20.7)	28.1 (± 17.5)	28.5 (± 18.8)	19.5 (± 16.2)	7.7 (± 16.8)	86.3 (± 16.4)	11.4 (± 18.5)	11.8 (± 13.7)
	GSD	6.3 (± 7.9)	7.1 (± 8.2)	28.9 (± 20.7)	26.8 (± 15.2)	26.9 (± 17.1)	21.0 (± 17.2)	10.0 (± 16.9)	86.1 (± 16.0)	10.4 (± 16.6)	10.0 (± 11.4)
	Lab	4.7 (± 8.1)	7.9 (± 10.9)	30.4 (± 22.3)	30.2 (± 18.0)	28.1 (± 18.1)	19.5 (± 15.4)	9.6 (± 18.6)	83.7 (± 18.3)	14.8 (± 20.6)	9.8 (± 11.2)
	Lab x GR	4.7 (± 11.9)	8.4 (± 11.4)	27.5 (± 21.9)	27.5 (± 16.9)	26.6 (± 16.4)	19.2 (± 14.9)	7.0 (± 15.3)	84.8 (± 17.5)	9.8 (± 15.4)	11.1 (± 12.6)
	Lab x GR* + Lab x Lab*	4.6 (± 7.7)	9.6 (± 14.6)	29.0 (± 20.8)	28.4 (± 15.8)	29.8 (± 17.7)	18.3 (± 14.2)	7.0 (± 14.5)	81.8 (± 20.8)	9.0 (± 15.7)	9.5 (± 9.2)
	SP x Lab + Lab x SP	4.6 (± 7.7)	7.8 (± 7.8)	23.3 (± 19.6)	24.4 (± 19.7)	22.6 (± 15.4)	14.8 (± 13.6)	11.5 (± 18.0)	90.6 (± 12.5)	8.6 (± 12.7)	9.9 (± 12.0)

Table 8.14 Influence of additional factors on 12M r-PWQ2 scores. Mean ($\pm$ standard deviation) for factor categories for components 1-10 at 12M. The lowest score for the factor is highlighted

Factor		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
PR experience	< 5	5.0 ($\pm$ 9.9)	10.0 ($\pm$ 13.8)	31.7 ($\pm$ 22.0)	30.7 ($\pm$ 17.5)	29.4 ($\pm$ 18.0)	19.8 ($\pm$ 15.7)	7.9 ($\pm$ 16.4)	83.8 ($\pm$ 18.2)	11.2 ($\pm$ 18.0)	10.7 ($\pm$ 12.2)
	5 to 9	4.7 ($\pm$ 9.8)	10.3 ($\pm$ 14.2)	26.1 ($\pm$ 20.1)	26.1 ($\pm$ 16.5)	26.9 ($\pm$ 17.9)	18.3 ($\pm$ 14.3)	7.1 ($\pm$ 15.3)	84.8 ($\pm$ 18.0)	10.2 ($\pm$ 15.6)	11.8 ($\pm$ 13.1)
	10 to 19	4.1 ($\pm$ 8.1)	9.2 ($\pm$ 11.7)	25.5 ($\pm$ 19.8)	27.0 ($\pm$ 16.2)	26.5 ($\pm$ 17.2)	17.7 ($\pm$ 14.6)	9.4 ($\pm$ 17.2)	82.3 ($\pm$ 20.2)	12.1 ($\pm$ 18.2)	10.5 ($\pm$ 12.1)
	20 +	4.3 ($\pm$ 8.5)	8.8 ($\pm$ 15.3)	24.5 ($\pm$ 20.2)	24.7 ($\pm$ 16.4)	22.2 ($\pm$ 18.0)	17.3 ($\pm$ 16.2)	9.5 ($\pm$ 15.8)	87.0 ($\pm$ 14.9)	10.7 ($\pm$ 18.5)	10.9 ($\pm$ 10.9)
Adults	1	6.1 ($\pm$ 10.8)	11.6 ($\pm$ 14.4)	28.4 ($\pm$ 21.6)	28.7 ($\pm$ 17.7)	30.2 ($\pm$ 17.4)	22.2 ($\pm$ 14.8)	8.1 ($\pm$ 14.8)	84.8 ($\pm$ 17.6)	10.5 ($\pm$ 17.3)	12.9 ($\pm$ 13.8)
	2	4.4 ($\pm$ 9.7)	9.1 ($\pm$ 21.1)	29.2 ($\pm$ 18.6)	29.0 ($\pm$ 16.8)	27.2 ($\pm$ 17.8)	18.6 ($\pm$ 15.2)	7.5 ($\pm$ 16.5)	84.9 ($\pm$ 17.5)	11.0 ($\pm$ 17.7)	9.9 ($\pm$ 11.3)
	3	4.6 ($\pm$ 7.6)	10.2 ($\pm$ 12.7)	29.8 ($\pm$ 21.4)	28.3 ($\pm$ 17.7)	28.7 ($\pm$ 19.3)	18.3 ($\pm$ 15.1)	8.9 ($\pm$ 15.9)	80.8 ($\pm$ 21.0)	11.9 ($\pm$ 17.5)	12.6 ($\pm$ 13.6)
	4	4.0 ($\pm$ 8.2)	12.2 ($\pm$ 16.7)	29.8 ($\pm$ 23.2)	27.9 ($\pm$ 17.6)	27.8 ($\pm$ 16.8)	16.5 ($\pm$ 15.2)	9.4 ($\pm$ 17.7)	82.1 ($\pm$ 19.0)	10.9 ($\pm$ 16.5)	11.0 ($\pm$ 11.7)
	5+	6.5 ($\pm$ 7.8)	10.6 ($\pm$ 17.9)	31.8 ($\pm$ 25.4)	29.5 ($\pm$ 18.3)	34.5 ($\pm$ 23.0)	22.2 ($\pm$ 21.2)	14.8 ($\pm$ 22.7)	76.4 ($\pm$ 24.2)	8.4 ($\pm$ 13.2)	17.3 ($\pm$ 19.8)
Children	Yes	6.4 ($\pm$ 10.8)	12.2 ($\pm$ 16.2)	33.2 ($\pm$ 22.4)	30.7 ($\pm$ 17.5)	31.5 ($\pm$ 19.1)	19.7 ($\pm$ 16.4)	10.6 ($\pm$ 19.1)	80.3 ($\pm$ 19.8)	12.1 ($\pm$ 17.9)	12.4 ($\pm$ 13.0)
	No	4.3 ($\pm$ 9.5)	10.1 ($\pm$ 14.1)	29.3 ($\pm$ 20.7)	29.1 ($\pm$ 17.2)	28.3 ($\pm$ 17.9)	18.9 ($\pm$ 15.2)	7.2 ($\pm$ 15.3)	84.4 ($\pm$ 17.3)	11.3 ($\pm$ 17.9)	11.1 ($\pm$ 12.2)
Dogs	Yes	5.3 ($\pm$ 10.1)	10.3 ($\pm$ 15.1)	28.5 ($\pm$ 21.9)	28.6 ($\pm$ 17.3)	28.6 ($\pm$ 19.3)	18.7 ($\pm$ 15.6)	10.8 ($\pm$ 19.2)	84.0 ($\pm$ 18.1)	10.9 ($\pm$ 17.6)	11.5 ($\pm$ 12.6)
	No	4.4 ($\pm$ 9.3)	9.8 ($\pm$ 13.1)	29.8 ($\pm$ 21.2)	29.2 ($\pm$ 17.2)	27.8 ($\pm$ 17.2)	19.2 ($\pm$ 15.3)	6.5 ($\pm$ 14.2)	83.9 ($\pm$ 18.5)	11.3 ($\pm$ 17.7)	10.7 ($\pm$ 12.3)
Animals	Yes	5.3 ($\pm$ 9.9)	10.4 ($\pm$ 14.5)	28.8 ($\pm$ 21.9)	28.8 ($\pm$ 17.3)	28.5 ($\pm$ 18.8)	18.9 ($\pm$ 15.7)	9.9 ($\pm$ 18.7)	83.8 ($\pm$ 18.4)	11.3 ($\pm$ 18.1)	11.3 ($\pm$ 12.5)

	No	4.1 (± 9.3)	9.5 (± 13.1)	29.7 (± 20.9)	29.2 (± 17.1)	27.8 (± 17.2)	19.1 (± 15.1)	6.2 (± 13.2)	84.1 (± 18.2)	11.0 (± 17.2)	10.6 (± 12.3)
Location	Rural	4.7 (± 10.2)	10.7 (± 15.6)	29.1 (± 20.7)	29.5 (± 17.2)	28.9 (± 19.1)	19.8 (± 16.3)	8.1 (± 16.6)	84.0 (± 19.3)	11.0 (± 18.0)	11.1 (± 13.0)
	Suburban	4.5 (± 8.8)	9.3 (± 12.5)	29.5 (± 21.7)	28.6 (± 17.0)	26.8 (± 17.0)	18.5 (± 14.7)	7.9 (± 16.5)	84.2 (± 17.8)	10.2 (± 16.4)	10.4 (± 11.6)
	Urban	5.8 (± 10.9)	10.4 (± 13.6)	28.3 (± 21.7)	28.7 (± 17.8)	30.6 (± 18.9)	18.9 (± 15.2)	8.2 (± 14.9)	83.3 (± 18.1)	14.1 (± 20.4)	12.5 (± 13.6)
Puppy Class	Yes	4.4 (± 8.2)	9.9 (± 13.5)	29.6 (± 21.8)	29.1 (± 17.6)	28.4 (± 18.4)	19.0 (± 15.1)	8.0 (± 16.0)	83.5 (± 18.6)	11.5 (± 17.6)	10.7 (± 11.9)
	No	7.2 (± 13.9)	8.7 (± 10.0)	27.7 (± 22.3)	28.6 (± 17.3)	26.1 (± 19.0)	17.8 (± 15.0)	8.5 (± 17.7)	85.1 (± 18.2)	8.7 (± 15.9)	9.0 (± 11.0)
Work	Yes	3.7 (± 7.9)	10.2 (± 11.5)	26.9 (± 21.1)	26.9 (± 17.1)	27.9 (± 19.2)	17.9 (± 15.5)	8.4 (± 17.2)	84.4 (± 16.8)	10.6 (± 15.9)	11.1 (± 12.8)
	No	4.9 (± 9.2)	9.7 (± 13.5)	29.8 (± 22.0)	29.5 (± 17.7)	28.1 (± 18.4)	19.0 (± 15.5)	7.9 (± 15.8)	83.7 (± 16.8)	11.4 (± 17.9)	10.4 (± 11.6)
YOB	2016	4.5 (± 11.5)	5.9 (± 7.1)	26.4 (± 19.2)	26.1 (± 20.4)	25.1 (± 18.8)	20.5 (± 19.8)	10.3 (± 20.4)	86.7 (± 17.1)	10.2 (± 16.5)	8.5 (± 12.9)
	2017	4.3 (± 8.3)	9.9 (± 14.2)	27.0 (± 20.4)	27.1 (± 16.9)	28.4 (± 18.1)	18.6 (± 15.2)	7.9 (± 16.3)	84.7 (± 17.7)	11.6 (± 18.9)	11.2 (± 13.0)
	2018	5.0 (± 10.4)	10.1 (± 13.5)	30.8 (± 22.2)	30.2 (± 17.0)	27.9 (± 17.7)	19.1 (± 15.2)	7.5 (± 15.2)	83.0 (± 19.0)	10.3 (± 15.9)	10.7 (± 11.5)
	2019	5.7 (± 11.5)	10.3 (± 13.2)	34.6 (± 21.5)	32.7 (± 17.9)	28.2 (± 18.6)	20.5 (± 15.1)	10.9 (± 19.9)	84.2 (± 18.1)	12.0 (± 18.2)	12.0 (± 12.6)
Neutered	Yes	4.9 (± 10.1)	9.7 (± 13.6)	28.8 (± 21.0)	28.8 (± 17.2)	28.2 (± 18.0)	19.8 (± 15.6)	7.6 (± 16.4)	84.2 (± 17.9)	11.1 (± 18.0)	11.0 (± 12.4)
	No	4.5 (± 8.9)	10.7 (± 14.2)	30.1 (± 22.1)	29.0 (± 17.0)	28.3 (± 18.0)	17.6 (± 14.7)	8.7 (± 16.2)	83.7 (± 18.8)	11.2 (± 16.9)	11.1 (± 12.5)

8.6 log likelihood comparisons for h² models

8.6.1 5M heritability models

Table 8.15 Model outputs for 5M C4 model 1, when including the following fixed factors: sex, proportion Labrador, Puppy raiser experience and presence of children.

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid	-1162.80	235.67	349	271.54	235.67	0.13
Pid, litter	-1162.80	235.67	349	271.54	235.67	0.13
Pid, PDA	-1162.80	235.67	349	271.54	235.67	0.13
Pid, PR ID	-1163.20	249.50	349	270.41	249.50	0.08
Pid, PDA, PR ID	-1163.31	248.15	349	269.28	248.15	0.08
Pid, litter, PDA	-1162.80	235.52	349	271.35	235.53	0.13
Pid, litter, PR ID	-1163.17	248.07	349	269.26	248.07	0.08
Pid, litter, PDA and PR ID	-1163.29	246.73	349	246.73	268.14	0.08

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

Then puppy walker experience was removed from the model and PR ID trialled as random factor

Table 8.16 Model output for 5M C4 model 2, when including the following fixed factors: sex, proportion Labrador and presence of children.

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid, PR ID	-1168.62	250.25	352	270.66	250.25	0.08
Pid, PTS, PR ID	-1168.78	248.97	352	269.61	248.97	0.08
Pid, litter, PR ID	-1168.62	248.85	352	269.53	248.86	0.08
Pid, litter, PDA, PR ID	-1168.79	247.59	352	268.49	247.59	0.08

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

8.6.2 8M heritability models

Table 8.17 Model outputs for 8M C 5 model 1, when including the following fixed factors: sex, proportion Labrador, home location, puppy raiser experience and presence of children.

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid	-4164.41	292.81	1226	320.43	292.81	0.09
Pid, litter	-4163.17	288.68	1226	319.17	288.68	0.05
Pid, PDA	-4163.45	289.00	1226	320.53	289.00	0.09
Pid, PR ID	-4145.12	180.40	1226	317.19	180.40	0.05
Pid, PDA, PR ID	-4144.75	179.49	1226	317.26	179.49	0.05
Pid, litter, PDA	-4162.40	285.68	1226	319.23	285.68	0.05
Pid, litter, PR ID	-4143.82	175.00	1226	316.42	175.00	0.02
Pid, litter, PDA and PR ID	-4143.55	174.42	1226	316.43	174.42	0.02

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

Then puppy walker experience was removed from the model and PW ID trialled as random factor

Table 8.18 Model outputs for 8M C5 model 2, when including the following fixed factors: sex, proportion Labrador, home location and presence of children

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid, PR ID	-4151.52	179.31	1230	197.43	179.31	0.09
Pid, PDA, PR ID	-4151.29	178.58	1230	317.23	178.58	0.06
Pid, litter, PR ID	-4150.00	173.65	1230	316.34	173.65	0.02
Pid, litter, PDA, PR ID	-4149.85	173.21	1230	316.37	173.21	0.04

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

8.6.3 12M heritability models

Table 8.19 Model outputs for 12M C 6 model 1, when including the following fixed factors: sex, proportion Labrador, number of adults in the home, puppy raiser experience, presence of children and age at neutering.

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid	-2896.34	217.87	888	235.73	217.87	0.08
Pid, litter	-2896.33	217.92	888	235.65	217.93	0.07
Pid, PDA	-2896.34	217.78	888	235.66	217.78	0.08
Pid, PR ID	-2884.51	115.15	888	236.72	115.15	0.11
Pid, PDA, PR ID	-2884.51	115.15	888	236.72	115.15	0.11
Pid, litter, PDA	-2896.33	217.85	888	235.60	217.85	0.07
Pid, litter, PR ID	-2884.51	115.56	888	236.58	115.56	0.11
Pid, litter, PDA and PR ID	-2884.51	115.51	888	236.52	115.51	0.11

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

Then puppy walker experience was removed from the model and PW ID trialled as random factor

Table 8.20 Model outputs for 12M C 6 model 2, when including the following fixed factors: sex, proportion Labrador, number of adults in the home, presence of children and age at neutering.

Random factors in model	loglikelihood	S2	df	Phenotypic variance	Residual variance	Pid h ²
Pid, PR ID	-2890.42	115.03	892	236.27	115.03	0.11
Pid, PDA, PR ID	-2890.42	115.03	892	236.26	115.03	0.11
Pid, litter, PR ID	-2890.42	115.05	892	236.29	115.05	0.11
Pid, litter, PDA, PR ID	-2890.42	115.04	892	236.29	115.04	0.11

* Pid (individual), PDA (Puppy Development Advisor), PR ID (Puppy Raiser ID)

8.7 r-PWQ2 fixed factor effects

Table 8.21 Fixed factor effects for 5M heritability outputs. Significant effects are highlighted in a blue square.

Fixed Factors	C1		C2		C3		C4		C5		C6		C7		C8		C9		C10	
	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.
Children	4.99	2.11	2.03	1.43	3.96	1.65	3.79	1.99	2.25	2.01	7.05	2.73	-1.87	2.02	3.74	2.07	3.31	2.13	5.15	1.66
PWexp 2	-7.49	2.33	-0.11	1.58	0.45	1.84	-3.46	2.21	0.98	2.21	-7.69	3.00	-1.33	2.24	1.03	2.28	2.41	2.36	-0.03	1.83
PWexp 3	-4.78	3.22	0.73	2.17	4.11	2.52	1.45	3.03	0.11	3.05	-2.27	4.15	0.55	3.08	0.48	3.14	0.56	3.25	1.38	2.53
PWexp 4	-11.19	4.87	-2.86	3.29	-2.20	3.82	-5.78	4.60	-12.45	4.62	-10.73	6.29	-8.57	4.67	-0.41	4.77	5.13	4.92	0.57	3.83
Prop Lab	0.74	3.20	-0.84	2.18	-7.24	4.81	1.14	4.22	1.14	2.33	2.53	3.17	3.76	4.25	2.24	2.40	0.16	4.45	-2.07	2.49
Sex	-2.41	1.86	0.94	1.26	-2.14	1.47	-1.88	1.76	-1.96	1.76	-2.08	2.40	-2.96	1.79	0.07	1.82	0.13	1.89	0.08	1.47

Table 8.22 Fixed factor effects for 8M heritability outputs. Significant effects are highlighted in a blue square.

Fixed Factors	C1		C2		C3		C4		C5		C6		C7		C8		C9		C10	
	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.
Children	1.12	1.30	1.78	0.89	1.24	0.78	0.04	1.11	2.92	1.13	3.06	1.47	-5.43	1.10	0.56	1.00	-0.75	1.14	3.52	1.10
PWexp 1	6.68	1.49	-1.04	1.03	1.63	0.90	2.33	1.28	-2.19	1.30	2.84	1.69	-1.28	1.26	-2.04	1.15	0.93	1.32	-2.08	1.26
PWexp 3	-1.67	2.19	1.97	1.51	0.48	1.32	-1.07	1.88	-1.76	1.90	0.06	2.48	-0.80	1.85	-1.03	1.69	-1.71	1.93	1.08	1.85
PWexp 4	-7.91	3.55	-4.46	2.45	-2.95	2.14	3.35	3.05	-3.57	3.08	-8.60	4.02	-3.23	3.01	6.47	2.74	5.02	3.13	0.86	3.00
Location 2	2.18	1.31	-0.81	0.91	1.57	0.79	-0.74	1.13	-0.48	1.14	-0.01	1.49	0.18	1.11	0.49	1.02	0.21	1.16	-0.66	1.11
Location 3	4.97	1.81	3.10	1.25	2.92	1.09	2.04	1.55	2.88	1.57	6.82	2.05	-1.57	1.53	3.26	1.40	3.15	1.59	3.44	1.53
Prop Lab	4.70	3.68	-6.35	1.68	0.57	1.94	0.44	2.02	4.43	3.35	6.84	3.61	5.06	2.52	0.43	2.02	3.37	2.31	1.37	1.52
Sex	-0.55	1.17	-1.18	0.80	1.39	0.70	-0.73	1.00	-0.02	1.02	2.83	1.32	-3.32	0.99	0.02	0.90	-0.37	1.03	1.45	0.98

Table 8.23 Fixed factor effects for 12M heritability outputs. Significant effects are highlighted in a blue square.

Fixed Factors	C1		C2		C3		C4		C5		C6		C7		C8		C9		C10	
	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.	E	S.E.
Children	2.10	0.78	3.17	1.10	2.39	1.65	0.94	1.35	3.16	1.41	0.86	1.20	2.75	1.30	-3.47	1.40	0.72	1.44	1.33	0.99
PWexp 1	0.39	0.86	-1.03	1.20	4.04	1.79	3.60	1.47	1.17	1.54	1.53	1.31	-0.27	1.42	-0.51	1.53	0.02	1.57	-1.77	1.08
PWexp 3	-0.98	1.22	-1.04	1.71	0.36	2.56	1.58	2.10	-1.56	2.20	0.15	1.87	-0.84	2.03	-2.35	2.18	0.95	2.24	-2.30	1.53
PWexp 4	1.57	2.25	-3.67	3.15	-2.04	4.74	0.43	3.89	-2.79	4.07	0.62	3.45	4.95	3.75	2.39	4.03	1.00	4.13	2.93	2.84
Prop Lab	-0.16	1.85	-6.46	2.85	0.41	2.41	-2.22	1.98	0.53	2.81	0.68	3.03	0.37	3.15	-2.72	3.39	6.29	3.66	-3.26	2.03
Sex	1.00	1.09	0.99	1.53	4.45	2.29	4.68	1.87	4.73	1.97	0.58	1.67	-0.05	1.82	-2.42	1.95	-2.68	2.00	1.29	1.37
Adult 1	1.84	3.25	6.10	4.55	2.69	6.84	3.85	5.61	-7.42	5.87	-0.21	4.98	-12.18	5.41	3.51	5.82	2.28	5.96	-7.39	4.10
Adult 2	-0.07	3.19	3.38	4.45	5.13	6.70	4.50	5.49	-9.85	5.74	-2.58	4.87	-12.84	5.30	4.24	5.70	3.36	5.83	-9.72	4.01
Adult 3	-0.62	3.28	5.24	4.58	5.43	6.90	3.97	5.66	-8.39	5.92	-2.95	5.02	-10.62	5.45	1.36	5.87	5.25	6.00	-6.69	4.13
Adult 4	-1.77	3.54	6.90	4.95	5.93	7.45	1.80	6.10	-10.33	6.38	-5.44	5.41	-11.86	5.89	1.00	6.33	2.35	6.48	-8.02	4.46
Neutering	0.01	0.01	0.02	0.01	0.03	0.01	0.03	0.01	0.04	0.01	0.00	0.01	0.02	0.01	-0.01	0.01	-0.01	0.01	0.01	0.01

8.8 Behavioural assessment data sheet

Pup ID: _____ Marking: _____ Litter: _____ Date: _____ E: _____ H: _____.

Session 1: Start time _____.

1. Unsolvable

*camera glasses

FAMILIARISATION				
Trial	Complete	Repeat		
F1				
F2				
F3				
F4				

TEST		
Trial	Time looking at face	Vocal?
1		
2		
3		
4		

Notes:

3. Odour Discrimination

* baited arms

FAMILIARISATION		
Trial	Hand	Complete?
1	R	
2	L	

TEST						
Trial	Food	First choice	Last choice	Switch between?	Vocal?	Repeat
1	R					
2	L					
3	L					
4	R					
5	L					
6	R					

Notes:

2. Human interest

*camera glasses

Trial	E speaking to puppy	Play break		
	Time Looking at E	Vocal?	Time close to E	Vocal?
1				
2				
3				

Notes:

4. Novel object

* don't forget camera

* Check cat works!

Notes:

End Time:

Session 2: Start time: _____.

KEY
 F = Front
 T = Top
 B = Back
 S = Side
 L = long way round
 Sh = short way round

5. Cylinder

Warmups					
Bowl Trials x2			Leading trials (R x2)		

1. FAMILIARISATION (black cylinder – block to left)									
Trial	Correct (1/0)	1 st Touch (F, T, B, S)	Time to solve	Repeats	Trial	Correct (1/0)	1 st Touch (F, T, B, S)	Time to solve	Repeats
F1					F11				
F2					F12				
F3					F13				
F4					F14				
F5					F15				
F6					F16				
F7					F17				
F8					F18				
F9					F19				
F10					F20				

N.B if cut off at 10mins – just record how far the puppy got in the time

2. INHIBITORY CONTROL (clear cylinder – block Left)							3. DETOUR (change to block Right)						
Trial	Correct (1/0)	1 st Touch F T B S	Time to solve	L / Sh way round	Vocal	R	Trial	Correct (1/0)	1 st Touch F T B S	Time to solve	L / Sh way round	Vocal	R
T1							D1						
T2							D2						
T3							D3						
T4							D4						
T5							D5						
T6							T6						
T7							D7						
T8							D8						

6. Surprising events

*don't forget camera!

Notes:

1. Bin bag

2. Umbrella

3. Metal sheet

End Time:

8.9 Equipment used for behavioural assessment

Images of a) unsolvable box; b) odour elbow c) interactive cat; d) clear cylinder; e) testing mat; f) opaque cylinder and g) pot used for training cylinder and unsolvable. Images of a bin bag, umbrella and metal sheet are not shown.



8.10 Behavioural measures summary

Table 8.24 Behavioural assessment measures. Minimum, maximum, mean, percentage of variance, number of subjects with missing data and inclusion in the final data set for all behavioural measures collected during the assessment.

Measure	Task	Include in final data set	Min	Max	Mean	% variation	Missing data
Total looking time	Unsolvable	X	0.00	28.98	2.83	31.6%	16
Average looking time	Unsolvable	✓	0.00	7.25	0.71	31.6%	16
Total manipulation time	Unsolvable	X	0.00	96.38	44.74	1.3%	16
Average manipulation time	Unsolvable	✓	0.00	24.10	11.19	1.3%	16
Number vocalised (trials)	Unsolvable	✓	0.00	4.00	0.52	67.1%	16
Maximum vocalisation	Unsolvable	✓	0.00	3.00	0.64	67.1%	16
Number on experimenter	Unsolvable	✓	0.00	4.00	0.89	49.1%	16
Total looking time (talking)	Human Interest	X	0.00	58.48	9.95	5.3%	0
Average looking time (talking)	Human Interest	✓	0.00	19.49	3.32	5.3%	0
Number vocalised (talking)	Human Interest	✓	0.00	3.00	1.78	19.5%	0
Maximum vocalisation (talking)	Human Interest	✓	0.00	4.00	1.85	19.5%	0
Orientated towards experimenter (talking)	Human Interest	✓	1.64	26.65	13.08	0.0%	0
Total interaction with experimenter	Human Interest	X	0.00	90.00	75.67	0.3%	0
Average interaction with experimenter	Human Interest	✓	0.00	30.00	25.22	0.3%	0
Number vocalised (interaction)	Human Interest	✓	0.00	3.00	0.42	70.2%	0
Maximum vocalisation (interaction)	Human Interest	✓	0.00	4.00	0.62	70.2%	0
First choice correct	Odour	✓	0.00	1.00	0.55	0.3%	6
Last choice correct	Odour	✓	0.17	1.00	0.79	0.8%	6

Proximity to correct elbow	Odour	✓	3.88	116.42	72.71	0.0%	6
Proximity to incorrect elbow	Odour	✓	0.00	73.95	23.94	1.3%	6
Number vocalised (trials)	Odour	X	0.00	4.00	0.15	91.1%	6
Maximum vocalisation	Odour	X	0.00	3.00	0.15	91.1%	6
All trials completed	Odour	X	0.00	1.00	0.98	98.0%	6
First approach	Novel	✓	1.00	120.00	8.82	0.3%	1
Proximity total time	Novel	✓	0.00	117.00	53.91	2.3%	1
Number of approaches	Novel	✓	0.00	10.00	3.28	6.0%	1
Contact total	Novel	✓	0.00	102.00	23.05	6.0%	1
Orientation total	Novel	✓	17.00	118.00	60.29	0.0%	1
Grab (Y/N)	Novel	✓	0.00	1.00	0.39	61.3%	1
Crouch (Y/N)	Novel	✓	0.00	1.00	0.49	50.5%	1
Play bow (Y/N)	Novel	✓	0.00	1.00	0.15	84.7%	1
Latency to vocalisation	Novel	✓	7.00	120.00	40.34	0.0%	1
Vocal towards cat (Y/N)	Novel	✓	0.00	1.00	0.86	14.3%	1
Vocal towards door (Y/N)	Novel	✓	0.00	1.00	0.69	30.9%	1
Groan (Y/N)	Novel	✓	0.00	1.00	0.48	52.3%	1
Whine (Y/N)	Novel	✓	0.00	1.00	0.89	11.3%	1
Bark/yelp (Y/N)	Novel	✓	0.00	1.00	0.33	66.6%	1
Howl (Y/N)	Novel	✓	0.00	1.00	0.20	79.9%	1
Maximum vocal intensity	Novel	✓	0.00	4.00	2.53	4.3%	1
Latency to contact	Novel	✓	1.00	120.00	14.68	0.0%	1
2 nd approach latency	Novel	✓	0.00	45.00	1.93	57.0%	1
2 nd approach score	Novel	✓	1.00	5.00	1.48	57.3%	1
2 nd approach lure	Novel	X	0.00	1.00	0.00	99.7%	1
3 rd approach latency	Novel	✓	1.00	45.00	3.81	0.0%	1
3 rd approach path	Novel	✓	1.00	4.00	2.41	0.0%	1
3 rd approach lure	Novel	X	0.00	1.00	0.00	99.7%	1
Bag response score	Surprising	✓	1.00	4.00	1.39	68.6%	29
Bag latency to approach	Surprising	✓	1.00	45.00	15.12	0.0%	29
Bag approach score	Surprising	✓	1.00	4.00	1.50	65.1%	29

Bag approach path	Surprising	✓	1.00	4.00	2.40	47.6%	29
Bag minimum distance score	Surprising	✓	1.00	5.00	1.66	59.2%	29
Bag latency to contact	Surprising	✓	1.00	45.00	13.53	0.0%	29
Bag contact time	Surprising	✓	0.00	14.00	5.50	33.5%	29
Bag vocalisation	Surprising	✓	0.00	1.00	0.16	84.3%	29
Bag maximum vocal intensity	Surprising	✓	0.00	3.00	0.28	84.6%	29
Bag latency to reapproach	Surprising	✓	1.00	45.00	4.48	0.0%	29
Bag reapproach path	Surprising	✓	1.00	4.00	2.11	27.3%	29
Bag reapproach minimum distance score	Surprising	X	1.00	3.00	1.06	95.4%	138
Umbrella response score	Surprising	✓	1.00	4.00	2.87	5.9%	29
Umbrella latency to approach	Surprising	✓	0.00	45.00	9.21	2.2%	29
Umbrella approach score	Surprising	X	1.00	5.00	1.22	86.2%	29
Umbrella approach path	Surprising	✓	1.00	4.00	2.02	50.8%	29
Umbrella minimum distance score	Surprising	X	1.00	5.00	1.13	91.1%	29
Umbrella latency to contact	Surprising	✓	0.00	45.00	6.33	0.5%	29
Umbrella contact time	Surprising	✓	0.00	14.00	6.21	14.1%	29
Umbrella vocalisation	Surprising	✓	0.00	1.00	0.16	84.3%	29
Umbrella maximum vocal intensity	Surprising	✓	1.00	3.00	0.30	83.8%	29
Umbrella latency to reapproach	Surprising	✓	1.00	45.00	3.71	2.4%	29
Umbrella reapproach path	Surprising	✓	1.00	4.00	1.97	63.1%	30
Umbrella reapproach minimum distance score	Surprising	X	1.00	5.00	1.12	96.6%	80
Sheet response score	Surprising	✓	1.00	5.00	2.98	16.5%	30

Sheet latency to approach	Surprising	✓	0.00	45.00	14.08	4.1%	30
Sheet approach score	Surprising	✓	1.00	4.00	1.55	52.3%	30
Sheet approach path	Surprising	✓	1.00	4.00	3.09	11.9%	30
Sheet minimum distance score	Surprising	X	1.00	5.00	1.16	93.2%	30
Sheet latency to contact	Surprising	✓	0.00	45.00	13.93	4.1%	30
Sheet contact time	Surprising	✓	0.00	11.00	1.40	59.9%	30
Sheet vocalisation	Surprising	✓	0.00	1.00	0.55	45.0%	30
Sheet maximum vocal intensity	Surprising	✓	0.00	3.00	1.04	45.0%	30
Sheet latency to reapproach	Surprising	✓	1.00	19.00	3.60	0.0%	30
Sheet reapproach path	Surprising	✓	1.00	5.00	1.92	48.5%	30
Sheet reapproach minimum distance score	Surprising	X	1.00	5.00	1.10	96.3%	100
Alone time – orientation to handlers	Surprising	✓	0.00	58.00	20.35	4.3%	31
Alone time latency to vocalisation	Surprising	✓	1.00	60.00	42.56	0.0%	32
Alone vocalisation	Surprising	✓	0.00	1.00	0.60	40.3%	32
Alone maximum vocal intensity	Surprising	✓	0.00	4.00	1.23	40.6%	32
Inhibitory score	Cylinder	✓	0.00	0.88	0.38	10.9%	108
Detour score	Cylinder	✓	0.00	0.88	0.17	36.6%	112
Average latency to solve – familiarisation	Cylinder	✓	1.75	30.00	6.79	0.0%	43
Average latency to solve – inhibitory	Cylinder	✓	1.82	20.99	4.55	0.0%	108
Average latency to solve – detour	Cylinder	✓	1.53	24.74	8.26	0.0%	112
Detour direct approach	Cylinder	✓	0.00	100.00	21.78	42.5%	112
Inhibitory vocalisation	Cylinder	X	0.00	6.00	0.07	96.2%	108
Detour vocalisation	Cylinder	X	0.00	4.00	0.11	93.0%	112

Number of familiarisation warm-ups	Cylinder	✓	4.00	20.00	8.51	0.0%	97
Aborted stage	Cylinder	X	1.00	3.00	1.30	82.7%	330

8.11 Behavioural measures heritability

Table 8.25 Heritabilities, sex, breed and age effects for behavioural measures. Heritabilities above 0.5 are highlighted in dark blue and have two asterisks, heritabilities between 0.3 and 0.5 are highlighted in a middle shade of blue with one asterisk and heritabilities between 0.2 and 0.3 are shaded light blue.

Measure	Task	Include in final data set	h^2	h^2 SE	Sex effects ?	Breed effects?	Age effects?
Total looking time	Unsolvable	X	0.04	0.07			
Average looking time	Unsolvable	✓	0.04	0.07			
Total manipulation time	Unsolvable	X	0.51**	0.13			
Average manipulation time	Unsolvable	✓	0.51**	0.13			
Number vocalised (trials)	Unsolvable	✓	0.16	0.09			✓
Maximum vocalisation	Unsolvable	✓	0.27	0.10			
Number on experimenter	Unsolvable	✓	0.19	0.10			
Total looking time (talking)	Human Interest	X	0.24	0.11			✓
Average looking time (talking)	Human Interest	✓	0.24	0.11			✓
Number vocalised (talking)	Human Interest	✓	0.33*	0.11			
Maximum vocalisation (talking)	Human Interest	✓	0.35*	0.12	✓		
Orientated towards experimenter (talking)	Human Interest	✓	0.55**	0.13			
Total interaction with experimenter	Human Interest	X	0.43*	0.13			
Average interaction with experimenter	Human Interest	✓	0.43*	0.13			
Number vocalised (interaction)	Human Interest	✓	0.24	0.10			
Maximum vocalisation (interaction)	Human Interest	✓	0.22	0.10			
First choice correct	Odour	✓	0.08	0.08			
Last choice correct	Odour	✓	0.30*	0.12			✓
Proximity to correct elbow	Odour	✓	0.54**	0.13			
Proximity to incorrect elbow	Odour	✓	0.34*	0.11			✓
Number vocalised (trials)	Odour	X	0.26	0.11	✓		
Maximum vocalisation	Odour	X	0.06	0.08			
All trials completed	Odour	X	0.09	0.09			
First approach	Novel	✓	0.02	0.06			
Proximity total time	Novel	✓	0.34*	0.12			

Number of approaches	Novel	✓	0.21	0.10			
Contact total	Novel	✓	0.62 ^{***}	0.13			
Orientation total	Novel	✓	0.34 [*]	0.12			
Grab (Y/N)	Novel	✓	0.41 [*]	0.12			
Crouch (Y/N)	Novel	✓	0.25	0.11			
Play bow (Y/N)	Novel	✓	0.60 ^{***}	0.13			
Latency to vocalisation	Novel	✓	0.35 [*]	0.12			
Vocal towards cat (Y/N)	Novel	✓	0.19	0.10			
Vocal towards door (Y/N)	Novel	✓	0.57 ^{***}	0.13			
Groan (Y/N)	Novel	✓	0.00	0.07		✓	
Whine (Y/N)	Novel	✓	0.23	0.10			
Bark/yelp (Y/N)	Novel	✓	0.22	0.11			
Howl (Y/N)	Novel	✓	0.14	0.10			
Maximum vocal intensity	Novel	✓	0.21	0.11			
Latency to contact	Novel	✓	0.04	0.07			
2 nd approach latency	Novel	✓	0.00	0.07			
2 nd approach score	Novel	✓	0.07	0.08			
2 nd approach lure	Novel	X	0.02	0.06			
3 rd approach latency	Novel	✓	0.01	0.07			
3 rd approach path	Novel	✓	0.23	0.11			
3 rd approach lure	Novel	X	0.00	0.00			
Bag response score	Surprising	✓	0.11	0.08			
Bag latency to approach	Surprising	✓	0.35 [*]	0.13			✓
Bag approach score	Surprising	✓	0.21	0.11			
Bag approach path	Surprising	✓	0.48 [*]	0.13			
Bag minimum distance score	Surprising	✓	0.83 ^{***}	0.14			
Bag latency to contact	Surprising	✓	0.30	0.12			✓
Bag contact time	Surprising	✓	0.47 [*]	0.13			
Bag vocalisation	Surprising	✓	0.15	0.10	✓		✓
Bag maximum vocal intensity	Surprising	✓	0.10	0.09			✓
Bag latency to reapproach	Surprising	✓	0.04	0.08			
Bag reapproach path	Surprising	✓	0.16	0.10	✓		
Bag reapproach minimum distance score	Surprising	X	NA	NA	NA	NA	NA
Umbrella response score	Surprising	✓	0.48 [*]	0.13			
Umbrella latency to approach	Surprising	✓	0.16	0.11			
Umbrella approach score	Surprising	X	0.09	0.10			

Umbrella approach path	Surprising	✓	0.16	0.10			
Umbrella minimum distance score	Surprising	X	0.42 [†]	0.14			
Umbrella latency to contact	Surprising	✓	0.05	0.09		✓	✓
Umbrella contact time	Surprising	✓	0.68 ^{***}	0.13	✓		
Umbrella vocalisation	Surprising	✓	0.24	0.11			✓
Umbrella maximum vocal intensity	Surprising	✓	0.22	0.11			✓
Umbrella latency to reapproach	Surprising	✓	0.20	0.11			
Umbrella reapproach path	Surprising	✓	0.20	0.12		✓	
Umbrella reapproach minimum distance score	Surprising	X	0.02	0.08		✓	✓
Sheet response score	Surprising	✓	0.58 ^{***}	0.14			
Sheet latency to approach	Surprising	✓	0.30 [†]	0.12			✓
Sheet approach score	Surprising	✓	0.22	0.11			✓
Sheet approach path	Surprising	✓	0.54 ^{***}	0.15			
Sheet minimum distance score	Surprising	X	0.19	0.10			✓
Sheet latency to contact	Surprising	✓	0.33 [†]	0.13			✓
Sheet contact time	Surprising	✓	0.40 [†]	0.14			✓
Sheet vocalisation	Surprising	✓	0.23	0.11			✓
Sheet maximum vocal intensity	Surprising	✓	0.24	0.12			✓
Sheet latency to reapproach	Surprising	✓	0.27	0.12	✓		✓
Sheet reapproach path	Surprising	✓	0.16	0.11		✓	
Sheet reapproach minimum distance score	Surprising	X	1.00 ^{***}	0.15			
Alone time – orientation to handlers	Surprising	✓	0.58 ^{***}	0.14			
Alone time latency to vocalisation	Surprising	✓	0.48 [†]	0.14			✓
Alone vocalisation	Surprising	✓	0.47 [†]	0.13			
Alone maximum vocal intensity	Surprising	✓	0.41 [†]	0.13			✓
Inhibitory score	Cylinder	✓	0.40 [†]	0.14			✓
Detour score	Cylinder	✓	0.33 [†]	0.13			✓
Average latency to solve – familiarisation	Cylinder	✓	0.49 [†]	0.14			✓
Average latency to solve – inhibitory	Cylinder	✓	0.18	0.14			✓

Average latency to solve – detour	Cylinder	✓	0.29	0.14			✓
Detour direct approach	Cylinder	✓	0.37 [*]	0.14			
Inhibitory vocalisation	Cylinder	X	0.02	0.08			
Detour vocalisation	Cylinder	X	0.06	0.10			✓
Number of familiarisation warm-ups	Cylinder	✓	0.31 [*]	0.13			✓
Aborted stage	Cylinder	X	NA	NA	NA	NA	NA

8.12 Behavioural assessment reliability models

Table 8.26 Reliability models for behavioural measures dependent on the type of data.

Behavioural measure	Task	Type of measure	Reliability models run
Average manipulation time	Unsolveable	Continuous	ICC, 2-way mixed, absolute agreement
Number vocalised (trials)	Unsolveable	Ordinal	Fleiss' Kappa
Maximum vocalisation	Unsolveable	Ordinal	Fleiss' Kappa
Number on experimenter	Unsolveable	Ordinal	Fleiss' Kappa
First choice correct	Odour	Continuous	ICC, 2-way mixed, absolute agreement
Last choice correct	Odour	Continuous	ICC, 2-way mixed, absolute agreement
Sum of switches	Odour	Ordinal	Fleiss' Kappa
First approach	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Proximity total time	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Number of approaches	Novel	Ordinal	Fleiss' Kappa
Contact total	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Orientation total	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Grab (Y/N)	Novel	Binary	Fleiss' Kappa
Crouch (Y/N)	Novel	Binary	Fleiss' Kappa
Play bow (Y/N)	Novel	Binary	Fleiss' Kappa
Latency to vocalisation	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Time vocalising	Novel	Continuous	ICC, 2-way mixed, absolute agreement
Vocal towards cat (Y/N)	Novel	Binary	Fleiss' Kappa
Vocal towards door (Y/N)	Novel	Binary	Fleiss' Kappa
Groan (Y/N)	Novel	Binary	Fleiss' Kappa
Whine (Y/N)	Novel	Binary	Fleiss' Kappa
Bark/yelp (Y/N)	Novel	Binary	Fleiss' Kappa
Howl (Y/N)	Novel	Binary	Fleiss' Kappa
Maximum vocal intensity	Novel	Ordinal	Fleiss' Kappa
Latency to contact	Novel	Continuous	ICC, 2-way mixed, absolute agreement
2nd approach score	Novel	Ordinal	Fleiss' Kappa
2nd approach lure	Novel	Binary	Fleiss' Kappa
3rd approach latency	Novel	Continuous	ICC, 2-way mixed, absolute agreement
3rd approach path	Novel	Ordinal	Fleiss' Kappa
Inhibitory score	Cylinder	Continuous	ICC, 2-way mixed, absolute agreement
Inhibitory direct approach	Cylinder	Ordinal	Fleiss' Kappa
Bag response score	Surprising	Ordinal	Fleiss' Kappa
Bag latency to approach	Surprising	Continuous	ICC, 2-way mixed, absolute agreement

Bag approach score	Surprising	Ordinal	Fleiss' Kappa
Bag approach path	Surprising	Ordinal	Fleiss' Kappa
Bag minimum distance score	Surprising	Ordinal	Fleiss' Kappa
Bag latency to contact	Surprising	Continuous	ICC, 2-way mixed, absolute agreement
Bag contact time	Surprising	Continuous	ICC, 2-way mixed, absolute agreement
Bag vocalisation	Surprising	Binary	Fleiss' Kappa
Bag maximum vocal intensity	Surprising	Ordinal	Fleiss' Kappa
Bag latency to reapproach	Surprising	Continuous	ICC, 2-way mixed, absolute agreement
Bag reapproach path	Surprising	Ordinal	Fleiss' Kappa
Bag reapproach minimum distance score	Surprising	Ordinal	Fleiss' Kappa

8.13 Genotype sample paperwork

Dear xxx

A project has recently been approved by Guide Dogs' Research Planning Group and aims to develop behavioural estimated breeding values (EBVs) for our breeding programme. As part of this project, we will be reviewing a large amount of historic behaviour information collected on dogs over the past three years from Puppy Behaviour Questionnaires and collecting saliva swabs to review genetic differences between related individuals.

We have therefore selected 150 dogs to have saliva swabs collected which will be sent off for genotyping later in the year. We are emailing you as [dog name and ID number] that you are currently responsible for on GDIR has been selected as part of the study. As the dog is qualified, we understand that under the current situation it may be more complex to collect a swab. As such we have tried to select guide dog owners that are suggested to have support at home on GDI-R and therefore will have support to collect a swab safely. We would like to know if you think the guide dog owner would be happy to be contacted to collect a non-invasive saliva swab from the dog (with support) and return it to us as the National Breeding Centre.

The guide dog owner will be provided with a swab, sampling instructions and a pre-paid envelope to return the swab. The format of the instructions is an A4 PDF document with pictures which we understand will not be accessible to guide dog owners. We can provide large print bullet point instructions and talk through the process to guide dog owners. We hope that by including guide dog owners with support, this process will be possible.

Please can you reply to this email to let us know if you think the guide dog | owner would be ok to be contacted regarding a swab, and if they have a preference on contact format.

If you are no longer responsible for the dog, please can you let us know the best person to contact.

Many thanks for your help with this study and please let me know if you have any questions.

Becky

Becky Hunt
Canine Science Associate

Figure 8.3 Genotype study participation email sent to a staff member requesting contact to a guide dog owner, volunteer or rehomer.



Dear **OWNER NAME**, |

A project is being undertaken at Guide Dogs that aims to develop behavioural estimated breeding values for our breeding programme. As part of this project, we will be reviewing a large amount of behaviour information collated over the past three years and collecting saliva swabs for a sample of dogs to review genetic differences between related individuals.

We would like your help with this study by collecting a non-invasive saliva swab for **DOG NAME**. You will find collection instructions and a collection swab kit enclosed in this pack. These are standard sampling kits which are regularly used by dog owners across the world. Please do not open the swab kit until you are ready to collect your sample and ensure that you do not touch the collection sponge to any other surface or material to prevent accidental collection of DNA that is not your dog's.

Collecting the sample is painless and we do not anticipate the saliva swabs causing your dog any distress. However, if your dog does show signs of not being happy with the sample collection, please stop. If you have any issues taking the sample, please let us know at canine.research@guidedogs.org.uk

Once you have collected your sample following the instructions, please label the collection tube with the dog's name and ID number and return the sample to the National Breeding Centre in the prepaid envelope provided. Collected samples can be stored at room temperature.

If you would like any further information or have any questions, please email canine.research@guidedogs.org.uk

Thank you for helping us and we hope you are enjoying your time with **DOG NAME**.

Becky Hunt
Canine Research Associate, Guide Dogs National Breeding Centre

Figure 8.4 Genotype study participation request letter sent to owners

Oral swab collection instructions

- Animal should not eat for 30 minutes or drink for 10 minutes before saliva collection.
- Do not scrape the animal's teeth or cheek with the sponge.
- Do not allow the animal to chew or bite the sponge.

Precautions

Tube contains liquid. Do NOT spill. This liquid preserves the DNA sample and is required by the lab to analyze your sample.

Attention: Do NOT remove the cap prior to sample collection.	
1 	• Place the sponge in animal's mouth at the cheek pouch as shown in figure. Collect saliva for 30 seconds by moving sponge and mopping saliva where it naturally pools (in the cheek pouch and under the tongue). • For animals older than 6 months, moderate restraint may be required.
Attention: Tube contains liquid, do NOT spill.	
2 	• Hold the tube upright and unscrew the cap from the tube.
3 	• Turn the cap upside down and place the oral swab in the tube. • Screw cap on tightly to prevent liquid sample from leaking during transport.
4 	• Invert and shake the tube vigorously 10 times to thoroughly mix sample.
5 	• Optional: using a permanent marker, clearly write the animal identification number on the white space available on the tube label.

Intended use: This product is designed for the collection of DNA from animals.

Warnings: Wash with water if the liquid comes in contact with eyes or skin. Do NOT ingest.

Store at room temperature 15–30°C

Material Safety Data Sheet (MSDS) available at www.dnagenotek.com

For Animal Use Only

Patent (www.dnagenotek.com/legal/notice)
PD-PRI-00347 Issue 1/2013-05
© 2013 DNA Genotek Inc. all rights reserved.

DNA Genotek Inc.
2 Beaverbrook Road
Ottawa, ON, Canada K2K 1L1
Subsidiary of OraSure Technologies, Inc.

Toll-free (North America): 1.866.813.6354
Tel: 613.723.5757 • Fax: 613.723.5057
www.dnagenotek.com
info@dnagenotek.com

DNAGENOTEK

www.dnagenotek.com

Figure 8.5 Genotype sample collection instructions sent to owners.

References

- Abdi, H. & Williams, L. J. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2, 433-459.
- Abramovs, N., Brass, A. & Tassabehji, M. 2020. Hardy-weinberg equilibrium in the large scale genomic sequencing era. *Front Genet*, 11, 210.
- Acharya, V., Fan, K. H., Snitz, B. E., Ganguli, M., DeKosky, S. T., et al. 2023. Meta-analysis of age-related cognitive decline reveals a novel locus for the attention domain and implicates a covid-19-related gene for global cognitive function. *Alzheimer's and Dementia*, 19, 5010-5022.
- Adewuyi, E. O., O'Brien, E. K., Nyholt, D. R., Porter, T. & Laws, S. M. 2022. A large-scale genome-wide cross-trait analysis reveals shared genetic architecture between alzheimer's disease and gastrointestinal tract disorders. *Communications biology*, 5, 691.
- ADUK. 2025. *About assistance dogs uk* [Online]. Available: <https://www.assistedogs.org.uk/> [Accessed 05 January 2025].
- AI, V. V. 2025. *Improving animal welfare with smarter ai* [Online]. Available: <https://www.vetvisionai.com/> [Accessed 25 January 2025].
- Åkerberg, H., Wilsson, E., Sallander, M., Hedhammar, Å., Lagerstedt, A.-S., et al. 2012. Test for personality characteristics in dogs used in research. *Journal of Veterinary Behavior*, 7, 327-338.
- Alberghina, D., Giunta, F., Gioe, M. & Panzera, M. 2020. Behavior test for seven-week old puppies (*canis familiaris*): Inter-rater reliability and factors associated with test performance. *PloS one*, 15, e0236271.
- Als, T. D., Kurki, M. I., Grove, J., Voloudakis, G., Therrien, K., et al. 2023. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nature medicine*, 29, 1832-1844.
- Amat, M., Le Brech, S., Camps, T. & Manteca, X. 2020. Separation-related problems in dogs: A critical review. *Advances in Small Animal Care*, 1, 1-8.
- Amfim, A., Bercea, L. C. & Cucu, N. 2024. Canine genetics and epidemiology of behavior in dogs. In: Claborn, D. M. (ed.) *Epizootics - outbreaks of animal disease*. Rijeka: IntechOpen.
- Amundsen, T., Sundstrøm, S., Buvik, T., Gederaas, O. A. & Haaverstad, R. 2014. Can dogs smell lung cancer? First study using exhaled breath and urine screening in unselected patients with suspected lung cancer. *Acta oncologica*, 53, 307-315.
- Andersson, L. 2001. Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics*, 2, 130-8.
- Aragam, N., Wang, K.-S. & Pan, Y. 2011. Genome-wide association analysis of gender differences in major depressive disorder in the netherlands nesda and ntr population-based samples. *Journal of affective disorders*, 133, 516-521.

- Arata, S., Momozawa, Y., Takeuchi, Y. & Mori, Y. 2010. Important behavioral traits for predicting guide dog qualification. *Journal of Veterinary Medical Science*, 72, 539-45.
- Arata, S., Takeuchi, Y., Inoue, M. & Mori, Y. 2014. "Reactivity to stimuli" is a temperamental factor contributing to canine aggression. *PloS one*, 9, e100767.
- Arden, R., Bensky, M. K. & Adams, M. J. 2016. A review of cognitive abilities in dogs, 1911 through 2016: More individual differences, please! *Current Directions in Psychological Science*, 25, 307-312.
- Aron, E. N., Aron, A. & Jagiellowicz, J. 2012. Sensory processing sensitivity: A review in the light of the evolution of biological responsivity. *Personality and Social Psychology Review*, 16, 262-282.
- Arvelius, P., Strandberg, E. & Fikse, W. F. 2014. The swedish armed forces temperament test gives information on genetic differences among dogs. *Journal of Veterinary Behavior-Clinical Applications and Research*, 9, 281-289.
- Arvidsson, L. K., Adriaensen, F., van Dongen, S., De Stobbeleere, N. & Matthysen, E. 2017. Exploration behaviour in a different light: Testing cross-context consistency of a common personality trait. *Animal Behaviour*, 123, 151-158.
- Asher, L., Blythe, S., Roberts, R., Toothill, L., Craigon, P. J., et al. 2013. A standardized behavior test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *Journal of Veterinary Behavior*, 8, 431-438.
- Asher, L., Harvey, N. D., Craigon, P. J., Blythe, S., Green, G., et al. 2016. The epidemiology of guide dog behaviour. Guide Dogs External Report.
- Asher, L., England, G. C. W., Sommerville, R. & Harvey, N. D. 2020. Teenage dogs? Evidence for adolescent-phase conflict behaviour and an association between attachment to humans and pubertal timing in the domestic dog. *Biol Lett*, 16, 20200097.
- Asp, H. E., Fikse, W. F., Nilsson, K. & Strandberg, E. 2015. Breed differences in everyday behaviour of dogs. *Applied Animal Behaviour Science*, 169, 69-77.
- Audrestch, H. M., Whelan, C. T., Grice, D., Asher, L., England, G. C., et al. 2015. Recognizing the value of assistance dogs in society. *Disability and health journal*, 8, 469-474.
- Axelsson, E., Ljungvall, I., Bhoumik, P., Conn, L. B., Muren, E., et al. 2021. The genetic consequences of dog breed formation—accumulation of deleterious genetic variation and fixation of mutations associated with myxomatous mitral valve disease in cavalier king charles spaniels. *PLoS Genetics*, 17, e1009726.
- Bai, J., Chen, S., Zhou, Z., Zhao, X., Song, M., et al. 2023. Association between th/maob gene single nucleotide polymorphisms and excitability in labrador retrievers. *Animal Genetics*, 54, 570-575.
- Balasso, P., Marchesini, G., Ughelini, N., Serva, L. & Andrighetto, I. 2021. Machine learning to detect posture and behavior in dairy cows: Information from an accelerometer on the animal's left flank. *Animals*, 11, 2972.

- Bálint, A., Rieger, G., Miklósi, Á. & Pongrácz, P. 2017. Assessment of owner-directed aggressive behavioural tendencies of dogs in situations of possession and manipulation. *Royal Society open science*, 4, 171040.
- Ball, R. D. 2013. Designing a gwas: Power, sample size, and data structure. *Genome-wide association studies and genomic prediction*. Springer.
- Barnard, S., Marshall-Pescini, S., Passalacqua, C., Beghelli, V., Capra, A., et al. 2016. Does subjective rating reflect behavioural coding? Personality in 2 month-old dog puppies: An open-field test and adjective-based questionnaire. *PloS one*, 11, e0149831.
- Barnard, S., Marshall-Pescini, S., Pelosi, A., Passalacqua, C., Prato-Previde, E., et al. 2017. Breed, sex, and litter effects in 2-month old puppies' behaviour in a standardised open-field test. *Scientific reports*, 7, 1802.
- Barnard, S., Kennedy, D., Watson, R., Valsecchi, P. & Arnott, G. 2019. Revisiting a previously validated temperament test in shelter dogs, including an examination of the use of fake model dogs to assess conspecific sociability. *Animals*, 9, 835.
- Bartlett, C. R. 1976. *Heritabilities and genetic correlations between hip dysplasia and temperament traits of seeing eye dogs*. Masters Thesis, Rutgers University. Department of Animal Sciences.
- Bassan, E., Mair, A., De Santis, M., Bugianelli, M., Loretto, E., et al. 2024. An overview of the literature on assistance dogs using text mining and topic analysis. *Frontiers in Veterinary Science*, 11, 1463332.
- Bateson, P. 2010. Independent inquiry into dog breeding. University of Cambridge.
- Batt, L. S., Batt, M. S., Baguley, J. A. & McGreevy, P. D. 2008. Factors associated with success in guide dog training. *Journal of Veterinary Behavior: Clinical Applications and Research*, 3, 143-151.
- Batt, L. S., Batt, M. S., Baguley, J. A. & McGreevy, P. D. 2009. The value of puppy raisers' assessments of potential guide dogs' behavioral tendencies and ability to graduate. *Anthrozoös*, 22, 71-76.
- Battaglia, C. L. 2009. Periods of early development and the effects of stimulation and social experiences in the canine. *Journal of Veterinary Behavior*, 4, 203-210.
- Bayne, T., Brainard, D., Byrne, R. W., Chittka, L., Clayton, N., et al. 2019. What is cognition? *Current Biology*, 29, R608-R615.
- BBC. 2008. *Pedigree dogs exposed* [Online]. Available: http://www.bbc.co.uk/pressoffice/pressreleases/stories/2008/08_august/19/dogs.shtml [Accessed 05 November 2020].
- Bel-Fenellós, C., Biencinto-López, C., Sáenz-Rico, B., Hernández, A., Sandoval-Talamantes, A. K., et al. 2023. Cognitive-behavioral profile in pediatric patients with syndrome 5p-; genotype-phenotype correlations. *Genes*, 14, 1628.
- Bellamy, K. K. L., Storengen, L. M., Handegård, K. W., Arnet, E. F., Prestrud, K. W., et al. 2018. Drd2 is associated with fear in some dog breeds. *Journal of Veterinary Behavior*, 27, 67-73.

- Belzung, C. & Philippot, P. 2007. Anxiety from a phylogenetic perspective: Is there a qualitative difference between human and animal anxiety? *Neural plasticity*, 2007, 059676.
- Bennett, P. C. & Rohlf, V. I. 2007. Owner-companion dog interactions: Relationships between demographic variables, potentially problematic behaviours, training engagement and shared activities. *Applied Animal Behaviour Science*, 102, 65-84.
- Bhowmik, N., Cook, S. R., Croney, C., Barnard, S., Romaniuk, A. C., et al. 2024. Heritability and genome-wide association study of dog behavioral phenotypes in a commercial breeding cohort. *Genes*, 15, 1611.
- Bjerknæs, M. B. 2019. *Genetic variation in candidate genes of behaviour and associations with behaviour phenotypes in dogs*. Veterinary Student Research Programme Research Thesis, Norwegian University of Life Sciences.
- Blackwell, E. J., Bradshaw, J. W. & Casey, R. A. 2013. Fear responses to noises in domestic dogs: Prevalence, risk factors and co-occurrence with other fear related behaviour. *Applied Animal Behaviour Science*, 145, 15-25.
- Bleuer-Elsner, S., Zamansky, A., Fux, A., Kaplun, D., Romanov, S., et al. 2019. Computational analysis of movement patterns of dogs with adhd-like behavior. *Animals*, 9, 1140.
- Blokland, G. A., Grove, J., Chen, C.-Y., Cotsapas, C., Tobet, S., et al. 2022. Sex-dependent shared and nonshared genetic architecture across mood and psychotic disorders. *Biol Psychiatry*, 91, 102-117.
- Blythe, S. 2020. *Factors influencing training success in guide dogs and the development of novel behavioural assessments*. PhD, University of Nottingham.
- Bohmanova, J., Miglior, F., Jamrozik, J., Van Doormaal, B. J., Hand, K. J., et al. 2010. Genetic analysis of return over feed in canadian holsteins. *Animal*, 4, 173-182.
- Boo, Y. J., Park, C. I., Kim, H. W., Kim, S. J. & Kang, J. I. 2019. Possible association of the ubiquitin-specific peptidase 46 gene (usp46) with affective temperamental traits in healthy korean volunteers. *Psychiatry investigation*, 16, 87.
- Borchers, M., Chang, Y., Proudfoot, K., Wadsworth, B., Stone, A., et al. 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *Journal of Dairy Science*, 100, 5664-5674.
- Bourdon, R. M. 2000. *Understanding animal breeding*, Prentice Hall Upper Saddle River, NJ.
- Bovenkerk, B. & Nijland, H. J. 2017. The pedigree dog breeding debate in ethics and practice: Beyond welfare arguments. *Journal of Agricultural and Environmental Ethics*, 30, 387-412.
- Bovo, S., Ribani, A., Muñoz, M., Alves, E., Araujo, J. P., et al. 2020. Whole-genome sequencing of european autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems. *Genetics Selection Evolution*, 52, 33.

- Boyko, A. R. 2011. The domestic dog: Man's best friend in the genomic era. *Genome Biology*, 12, 216.
- Bradshaw, J. & Goodwin, D. 1999. Determination of behavioural traits of pure-bred dogs using factor analysis and cluster analysis; a comparison of studies in the USA and uk. *Research in Veterinary Science*, 66, 73-76.
- Brady, K., Cracknell, N., Zulch, H. & Mills, D. S. 2018. A systematic review of the reliability and validity of behavioural tests used to assess behavioural characteristics important in working dogs. *Frontiers in Veterinary Science*, 5, 103.
- Braem, M., Asher, L., Furrer, S., Lechner, I., Würbel, H., et al. 2017. Development of the “highly sensitive dog” questionnaire to evaluate the personality dimension “sensory processing sensitivity” in dogs. *PloS one*, 12, e0177616.
- Bräuer, J., Hanus, D., Pika, S., Gray, R. & Uomini, N. 2020. Old and new approaches to animal cognition: There is not “one cognition”. *Journal of Intelligence*, 8, 28.
- Bray, E. E., MacLean, E. L. & Hare, B. A. 2015. Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognition*, 18, 1317-1329.
- Bray, E. E., Sammel, M. D., Cheney, D. L., Serpell, J. A. & Seyfarth, R. M. 2017. Effects of maternal investment, temperament, and cognition on guide dog success. *Proceedings of the National Academy of Sciences*, 114, 9128-9133.
- Bray, E. E., Levy, K. M., Kennedy, B. S., Duffy, D. L., Serpell, J. A., et al. 2019. Predictive models of assistance dog training outcomes using the canine behavioral assessment and research questionnaire and a standardized temperament evaluation. *Frontiers in Veterinary Science*, 49.
- Bray, E. E., Gruen, M. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., et al. 2020. Cognitive characteristics of 8-to 10-week-old assistance dog puppies. *Animal Behaviour*, 166, 193-206.
- Bray, E. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., Kennedy, B. S., et al. 2021a. Early-emerging and highly heritable sensitivity to human communication in dogs. *Current Biology*, 31, 3132-3136. e5.
- Bray, E. E., Gruen, M. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., et al. 2021b. Dog cognitive development: A longitudinal study across the first 2 years of life. *Animal Cognition*, 24, 311-328.
- Bray, E. E., Otto, C. M., Udell, M. A. R., Hall, N. J., Johnston, A. M., et al. 2021c. Enhancing the selection and performance of working dogs. *Frontiers in Veterinary Science*, 8.
- Briffa, M. & Weiss, A. 2010. Animal personality. *Current Biology*, 20, R912-4.
- Bryson, G. T., O'Neill, D. G., Brand, C. L., Belshaw, Z. & Packer, R. M. 2024. The doodle dilemma: How the physical health of ‘designer-crossbreed’ cockapoo, labradoodle and cavapoo dogs’ compares to their purebred progenitor breeds. *PloS one*, 19, e0306350.
- Budaev, S. V. 2010. Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, 116, 472-480.

- Burnett, E., Brand, C. L., O'Neill, D. G., Pegram, C. L., Belshaw, Z., et al. 2022. How much is that doodle in the window? Exploring motivations and behaviours of uk owners acquiring designer crossbreed dogs (2019-2020). *Canine Medicine and Genetics*, 9, 8.
- Buss, D. M. 1995. Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, 6, 1-30.
- Byosiére, S.-E. & Prasad-Shreckengast, S. 2020. Canine cognition. In: Vonk, J. & Shackelford, T. (eds.) *Encyclopedia of animal cognition and behavior*. Cham: Springer International Publishing.
- Calleja-Rodriguez, A., Pan, J., Funda, T., Chen, Z., Baison, J., et al. 2020. Evaluation of the efficiency of genomic versus pedigree predictions for growth and wood quality traits in scots pine. *BMC Genomics*, 21, 796.
- Calus, M. P. 2010. Genomic breeding value prediction: Methods and procedures. *Animal*, 4, 157-64.
- Cao, X., Liu, W.-P., Cheng, L.-G., Li, H.-J., Wu, H., et al. 2021. Whole genome analyses reveal significant convergence in obsessive-compulsive disorder between humans and dogs. *Science Bulletin*, 66, 187-196.
- Caron-Lormier, G., Harvey, N. D., England, G. C. & Asher, L. 2016. Using the incidence and impact of behavioural conditions in guide dogs to investigate patterns in undesirable behaviour in dogs. *Scientific reports*, 6, 23860.
- Cathomas, F., Stegen, M., Sigrist, H., Schmid, L., Seifritz, E., et al. 2015. Altered emotionality and neuronal excitability in mice lacking *kctd12*, an auxiliary subunit of gabab receptors associated with mood disorders. *Transl Psychiatry*, 5, e510.
- Cauchoux, M., Chow, P., Van Horik, J., Atance, C., Barbeau, E., et al. 2018. The repeatability of cognitive performance: A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170281.
- Chaloupková, H., Svobodová, I., Vápeník, P. & Bartoš, L. 2018. Increased resistance to sudden noise by audio stimulation during early ontogeny in german shepherd puppies. *PloS one*, 13, e0196553.
- Chambers, R. D., Yoder, N. C., Carson, A. B., Junge, C., Allen, D. E., et al. 2021. Deep learning classification of canine behavior using a single collar-mounted accelerometer: Real-world validation. *Animals* 11, 6 (2021).
- Chen, F., Zimmermann, M., Hekman, J., Lord, K., Logan, B., et al. 2021a. Advancing genetic selection and behavioral genomics of working dogs through collaborative science. *Frontiers in Veterinary Science*, 8.
- Chen, Z., Bian, C., Liao, K., Bishop, C. & Li, Y. 2021b. Validity and reliability of a phone app and stopwatch for the measurement of 505 change of direction performance: A test-retest study design. *Frontiers in Physiology*, 12, 743800.
- Cho, D., O'Berry, K., Possa-Paranhos, I. C., Butts, J., Palanikumar, N., et al. 2023. Paraventricular thalamic mc3r circuits link energy homeostasis with anxiety-related behavior. *Journal of Neuroscience*, 43, 6280-6296.

- Choi, J., Kim, H. B., Kang, Y., Kim, M., Lee, S. i., et al. 2017. Personality consistency analysis in cloned quarantine dog candidates. *Acta Veterinaria Brno*, 86, 273-283.
- Chopik, W. J. & Weaver, J. R. 2019. Old dog, new tricks: Age differences in dog personality traits, associations with human personality traits, and links to important outcomes. *Journal of Research in Personality*, 79, 94-108.
- Clarke, T.-K., Crist, R. C., Doyle, G. A., Weiss, A. R., Brandt, H., et al. 2014. Characterization of genetic variation in the *vgl4* gene in anorexia nervosa. *Psychiatric genetics*, 24, 183-184.
- Clay, L., Paterson, M. B., Bennett, P., Perry, G. & Phillips, C. C. 2020. Comparison of canine behaviour scored using a shelter behaviour assessment and an owner completed questionnaire, c-barq. *Animals*, 10, 1797.
- Cleghern, Z., Foster, M., Mealin, S., Williams, E., Holder, T., et al. 2021. A cloud data collection platform for canine behavioural prediction using objective sensor data. *International Journal of Cloud Computing*, 10, 247-264.
- ClinCalc. 2025. *ClinCalc llc* [Online]. Available: <https://clincalc.com/stats/power.aspx> [Accessed 16 April 2025].
- Cline, A. 2012. The genetics of deafness in dalmatians. *Microreviews in Cell and Molecular Biology*, 1.
- Cloninger, C. R. 1994. Temperament and personality. *Current opinion in neurobiology*, 4, 266-273.
- Cobb, M., Branson, N., McGreevy, P., Lill, A. & Bennett, P. 2015. The advent of canine performance science: Offering a sustainable future for working dogs. *Behavioural processes*, 110, 96-104.
- Cole, J. B., Franke, D. E. & Leighton, E. A. 2004. Population structure of a colony of dog guides. *Journal of Animal Science*, 82, 2906-2912.
- Congdon, J. V., Hosseini, M., Gading, E. F., Masousi, M., Franke, M., et al. 2022. The future of artificial intelligence in monitoring animal identification, health, and behaviour. *Animals*, 12, 1711.
- Cooper, E., Zulch, H. E. & Mills, D. S. 2025. Demographic factors predicting chasing behaviour in dogs. *Available at SSRN 4947119*.
- Coppinger, R. & Coppinger, L. 2001. *A new understanding of canine origin, behaviour and evolution*, Chicago, USA, University of Chicago Press.
- Cornelis, M. C., Rimm, E. B., Curhan, G. C., Kraft, P., Hunter, D. J., et al. 2014. Obesity susceptibility loci and uncontrolled eating, emotional eating and cognitive restraint behaviors in men and women. *Obesity*, 22, E135-E141.
- Craigon, P. J., Hobson-West, P., England, G. C., Whelan, C., Lethbridge, E., et al. 2017. "She's a dog at the end of the day": Guide dog owners' perspectives on the behaviour of their guide dog. *PloS one*, 12, e0176018.
- Csibra, B., Bunford, N. & Gácsi, M. 2024. Development of a human-analogue, 3-symptom domain dog adhd and functionality rating scale (dafrs). *Scientific reports*, 14, 1808.
- Cummins, R. A. & Gullone, E. Why we should not use the 5-point likert scales: The case for subjective quality of life measurement.

- Second International Conference on Quality of Life in Cities, 2000 Singapore. 77-93.
- Curb, L. A., Abramson, C. I., Grice, J. W. & Kennison, S. M. 2013. The relationship between personality match and pet satisfaction among dog owners. *Anthrozoös*, 26, 395-404.
- Czerwinski, V., McArthur, M., Smith, B., Hynd, P. & Hazel, S. 2016. Selection of breeding stock among australian purebred dog breeders, with particular emphasis on the dam. *Animals*, 6, 75.
- D'Eath, R., Conington, J., Lawrence, A., Olsson, I. & Sandøe, P. 2010. Breeding for behavioural change in farm animals: Practical, economic and ethical considerations. *Animal Welfare*, 19, 17-27.
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., et al. 2018. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature communications*, 9, 2098.
- Dawson, D., Chapman, J. & Thomas, M. J. 2012. Fatigue-proofing: A new approach to reducing fatigue-related risk using the principles of error management. *Sleep medicine reviews*, 16, 167-175.
- De Fruyt, F., Van de Wiele, L. & Van Heeringen, C. 2000. Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Personality and individual differences*, 29, 441-452.
- De Meester, R. H., Pluijmakers, J., Vermeire, S. & Laevens, H. 2011. The use of the socially acceptable behavior test in the study of temperament of dogs. *Journal of Veterinary Behavior*, 6, 211-224.
- De Palma, C., Viggiano, E., Barillari, E., Palme, R., Dufour, A., B., et al. 2005. Evaluating the temperament in shelter dogs. *Behaviour*, 142, 1307-1328.
- Deng, S. L., Hu, Z. L., Mao, L., Gao, B., Yang, Q., et al. 2021. The effects of kctd12, an auxiliary subunit of gaba(b) receptor in dentate gyrus on behavioral response to chronic social defeat stress in mice. *Pharmacol Res*, 163, 105355.
- Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., et al. 2003. David: Database for annotation, visualization, and integrated discovery. *Genome Biology*, 4, 1-11.
- Deshpande, G., Zhao, S., Waggoner, P., Beyers, R., Morrison, E., et al. 2024. Two separate brain networks for predicting trainability and tracking training-related plasticity in working dogs. *Animals*, 14, 1082.
- Dictionary, O. E. 2024a. Dimension, n., sense 2.A. *Oxford English Dictionary*. Oxford University Press.
- Dictionary, O. E. 2024b. Observation, n., sense ii.5.A. *Oxford English Dictionary*. Oxford University Press.
- Diederich, C. & Giffroy, J.-M. 2006. Behavioural testing in dogs: A review of methodology in search for standardisation. *Applied Animal Behaviour Science*, 97, 51-72.
- Diederich, C. 2024. The beauty of the beast: Suggestions to curb the excesses of dog breeding and restore animal welfare—invited review. *Veterinární medicína*, 69, 369.

- Diener, E. & Lucas, R. E. 2019. Personality traits. *General psychology: Required reading*, 278.
- Diesel, G., Brodbelt, D. & Pfeiffer, D. U. 2008. Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the uk. *Applied Animal Behaviour Science*, 115, 171-181.
- Dinwoodie, I. R., Dwyer, B., Zottola, V., Gleason, D. & Dodman, N. H. 2019. Demographics and comorbidity of behavior problems in dogs. *Journal of Veterinary Behavior*, 32, 62-71.
- Dodgson, J. B., Cheng, H. H. & Okimoto, R. 1997. DNA marker technology: A revolution in animal genetics. *Poultry Science*, 76, 1108-1114.
- Dogs, G. 2020. *The history of guide dogs* [Online]. Available: <https://www.guidedogs.org.uk/about-us/what-we-do/the-history-of-guide-dogs> [Accessed 24 August 2020].
- Dollion, N., Paulus, A., Champagne, N., St-Pierre, N., St-Pierre, É., et al. 2019. Fear/reactivity in working dogs: An analysis of 37 years of behavioural data from the mira foundation's future service dogs. *Applied Animal Behaviour Science*, 221, 104864.
- Döring, D., Roscher, A., Scheipl, F., Küchenhoff, H. & Erhard, M. H. 2009. Fear-related behaviour of dogs in veterinary practice. *The Veterinary Journal*, 182, 38-43.
- Douhovnikoff, V. & Leventhal, M. 2016. The use of hardy-weinberg equilibrium in clonal plant systems. *Ecol Evol*, 6, 1173-80.
- Dreger, D. L., Rimbault, M., Davis, B. W., Bhatnagar, A., Parker, H. G., et al. 2016. Whole-genome sequence, snp chips and pedigree structure: Building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. *Disease Models and Mechanisms*, 9, 1445-1460.
- Dreschel, N. A. 2010. The effects of fear and anxiety on health and lifespan in pet dogs. *Applied Animal Behaviour Science*, 125, 157-162.
- Duffy, D. L. & Serpell, J. A. 2008. Behavioral assessment of guide and service dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, 4, 186-188.
- Duffy, D. L. & Serpell, J. A. 2012. Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science*, 138, 99-109.
- Dunn, C. W. & Ryan, J. F. 2015. The evolution of animal genomes. *Current Opinion in Genetics & Development*, 35, 25-32.
- Einheber, S., Meng, X., Rubin, M., Lam, I., Mohandas, N., et al. 2013. The 4.1 b cytoskeletal protein regulates the domain organization and sheath thickness of myelinated axons. *Glia*, 61, 240-253.
- Eissman, J. M., Archer, D. B., Mukherjee, S., Lee, M. L., Choi, S. E., et al. 2024. Sex-specific genetic architecture of late-life memory performance. *Alzheimer's and Dementia*, 20, 1250-1267.
- Emery, N. J. & Clayton, N. S. 2009. Tool use and physical cognition in birds and mammals. *Current opinion in neurobiology*, 19, 27-33.
- Emily, E. C., Zulch, H. & Mills, D. S. 2024. The role of breed versus personality and other demographic factors in predicting chasing behaviour in dogs. *Applied Animal Behaviour Science*, 106463.

- Ennik, I., Liinamo, A.-E., Leighton, E. & van Arendonk, J. 2006. Suitability for field service in 4 breeds of guide dogs. *Journal of Veterinary Behavior*, 1, 67-74.
- Evans, K. M. 2014. *Genetic evaluation of guide dogs in the uk*. PhD, University of Nottingham.
- Evans, K. M., Lewis, T. W., Asher, L., Blythe, S., Bottomley, M., et al. 2015. Genetic evaluation of traits in a standardized behavioral test for potential guide dog puppies using crossbreed models. *Journal of Veterinary Behavior*, 10, 459-464.
- Evans, L., Cameron-Whytock, H. & Ijichi, C. 2024. Whoa, no-go: Evidence consistent with model-based strategy use in horses during an inhibitory task. *Applied Animal Behaviour Science*, 277, 106339.
- Eyre, A. W., Zapata, I., Hare, E., Serpell, J. A., Otto, C. M., et al. 2023. Machine learning prediction and classification of behavioral selection in a canine olfactory detection program. *Scientific reports*, 13, 12489.
- Falconer, D. S. & Mackay, T. F. C. 1989. *Introduction to qualitative genetics*. , Essex, England, Longman Scientific & Technical.
- Fallahi, M., Masoudi, A. A. & Torshizi, R. V. 2024. Genetic variants in the *tac1* transcriptional regulatory region affect on trainability and excitability levels in belgian malinois dogs. *Veterinary Medicine and Science*, 10, e1346.
- Fassino, S., Abbate-Daga, G., Amianto, F., Leombruni, P., Boggio, S., et al. 2002. Temperament and character profile of eating disorders: A controlled study with the temperament and character inventory. *International Journal of Eating Disorders*, 32, 412-425.
- Feng, L. C., Bullis, J. M., Kret, S., Famula, T. R. & Byosiére, S.-E. 2024. Evaluating the generalizability of a standardized behavior test in predicting programmatic success of guide dog puppies. *Applied Animal Behaviour Science*, 274, 106242.
- Fidler, A. E., van Oers, K., Drent, P. J., Kuhn, S., Mueller, J. C., et al. 2007. *Drd4* gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings of the Royal Society B*, 274, 1685-91.
- Fjukstad, K. K., Athanasiu, L., Bahrami, S., O'Connell, K. S., van der Meer, D., et al. 2021. Genetic variants associated with cardiometabolic abnormalities during treatment with selective serotonin reuptake inhibitors: A genome-wide association study. *The pharmacogenomics journal*, 21, 574-585.
- Flint, H. E., Coe, J. B., Pearl, D. L., Serpell, J. A. & Niel, L. 2018a. Effect of training for dog fear identification on dog owner ratings of fear in familiar and unfamiliar dogs. *Applied Animal Behaviour Science*, 208, 66-74.
- Flint, H. E., Coe, J. B., Serpell, J. A., Pearl, D. L. & Niel, L. 2018b. Identification of fear behaviors shown by puppies in response to nonsocial stimuli. *Journal of Veterinary Behavior*, 28, 17-24.
- Foltin, S. & Glenk, L. M. 2023. Current perspectives on the challenges of implementing assistance dogs in human mental health care. *Veterinary Sciences*, 10, 62.

- Fongang, B., Impairment, M. V. C., Dementia consortium, Sargurupremraj, M., Jian, X., et al. 2024. A genome-wide association meta-analysis of all-cause and vascular dementia. *Alzheimer's and Dementia*, 20, 5973-5995.
- Foraita, M., Howell, T. & Bennett, P. 2021. Working memory and inhibition in 7.5-week-old seeing eye dog puppies. *Applied Animal Behaviour Science*, 241, 105394.
- Foraita, M., Howell, T. & Bennett, P. 2024. Executive functions in a population of guide dog candidates. *Applied Animal Behaviour Science*, 270, 106126.
- Forman, O. P., Penderis, J., Hartley, C., Hayward, L. J., Ricketts, S. L., et al. 2012. Parallel mapping and simultaneous sequencing reveals deletions in bcan and fam83h associated with discrete inherited disorders in a domestic dog breed. *PLoS Genetics*, 8, e1002462.
- Forsyth, K. K., McCoy, B. M., Schmid, S. M., Promislow, D. E., Snyder-Mackler, N., et al. 2023. Lifetime prevalence of owner-reported medical conditions in the 25 most common dog breeds in the dog aging project pack. *Frontiers in Veterinary Science*, 10, 1140417.
- Fox, M. & Spencer, J. 1969. Exploratory behavior in the dog: Experiential or age dependent? *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 2, 68-74.
- Foyer, P., Bjällerhag, N., Wilsson, E. & Jensen, P. 2014. Behaviour and experiences of dogs during the first year of life predict the outcome in a later temperament test. *Applied Animal Behaviour Science*, 155, 93-100.
- Fratkin, J. L., Sinn, D. L., Thomas, S., Hilliard, S., Olson, Z., et al. 2015. Do you see what i see? Can non-experts with minimal training reproduce expert ratings in behavioral assessments of working dogs? *Behavioural processes*, 110, 105-116.
- Freedman, D. G., King, J. A. & Elliot, O. 1961. Critical period in the social development of dogs. *Science*, 133, 1016-1017.
- Friedrich, J., Strandberg, E., Arvelius, P., Sánchez-Molano, E., Pong-Wong, R., et al. 2019. Genetic dissection of complex behaviour traits in german shepherd dogs. *Heredity*, 123, 746-758.
- Funk, J. A. & Williams, W. L. 2020. Training the guide dog: An untapped opportunity for the behavior analyst. *Behavior Analysis: Research and Practice*, 20, 120-131.
- Garamszegi, L. Z., Eens, M. & Török, J. 2008. Birds reveal their personality when singing. *PloS one*, 3, e2647.
- Garcia, D., Lester, N., Cloninger, K. M. & Robert Cloninger, C. 2017. Temperament and character inventory (tci). In: Zeigler-Hill, V. & Shackelford, T. K. (eds.) *Encyclopedia of personality and individual differences*. Cham: Springer International Publishing.
- Gazit, I. & Terkel, J. 2003. Domination of olfaction over vision in explosives detection by dogs. *Applied Animal Behaviour Science*, 82, 65-73.
- Gazzano, A., Mariti, C., Notari, L., Sighieri, C. & McBride, E. A. 2008. Effects of early gentling and early environment on emotional

- development of puppies. *Applied Animal Behaviour Science*, 110, 294-304.
- Gelernter, J. & Polimanti, R. 2021. Genetics of substance use disorders in the era of big data. *Nature Reviews Genetics*, 22, 712-729.
- Ghirlanda, S., Acerbi, A., Herzog, H. & Serpell, J. A. 2013. Fashion vs. Function in cultural evolution: The case of dog breed popularity. *PloS one*, 8, e74770.
- Gialluisi, A., Andlauer, T. F., Mirza-Schreiber, N., Moll, K., Becker, J., et al. 2019. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. *Translational Psychiatry*, 9, 77.
- Giannakopoulou, O., Lin, K., Meng, X., Su, M.-H., Kuo, P.-H., et al. 2021. The genetic architecture of depression in individuals of east asian ancestry: A genome-wide association study. *JAMA psychiatry*, 78, 1258-1269.
- Gibson, G. 2018. Population genetics and gwas: A primer. *PLOS Biology*, 16, e2005485.
- Gibson, P. E. & Oliva, J. L. 2022. Public perceptions of australian assistance dogs: Happier and better used than companion dogs. *Journal of Applied Animal Welfare Science*, 25, 18-30.
- Ginja, M., Silvestre, A., Ferreira, A., Gonzalo-Orden, J., Orden, M., et al. 2008. Passive hip laxity in estrela mountain dog—distraction index, heritability and breeding values. *Acta Veterinaria Hungarica*, 56, 303-312.
- Goddard, M. E. & Beilharz, R. G. 1982. Genetic and environmental factors affecting the suitability of dogs as guide dogs for the blind. *Theoretical and Applied genetics*, 62, 97-102.
- Goddard, M. E. & Beilharz, R. G. 1983. Genetics of traits which determine the suitability of dogs as guide-dogs for the blind. *Applied Animal Ethology*, 9, 299-315.
- Goddard, M. E. & Beilharz, R. G. 1984. A factor analysis of fearfulness in potential guide dogs. *Applied Animal Behaviour Science*, 12, 253-265.
- Goddard, M. E. & Beilharz, R. G. 1986. Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*, 15, 247-260.
- Goes, F. S., McGrath, J., Avramopoulos, D., Wolyniec, P., Pirooznia, M., et al. 2015. Genome-wide association study of schizophrenia in ashkenazi jews. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168, 649-659.
- Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., et al. 1987. Roundtable: What is temperament? Four approaches. *Child development*, 58, 505-29.
- Goldsmith, H. H. & Lemery, K. S. 2000. Linking temperamental fearfulness and anxiety symptoms: A behavior–genetic perspective. *Biol Psychiatry*, 48, 1199-1209.
- González-Martínez, Á., Muñiz de Miguel, S. & Diéguez, F. J. 2024. New advances in attention-deficit/hyperactivity disorder-like dogs. *Animals*, 14, 2067.

- González-Ramírez, M. T., Quezada-Berumen, L. & Landero-Hernández, R. 2017. Assessment of canine behaviors using c-barq in a sample from northern Mexico. *Journal of Veterinary Behavior*, 20, 52-58.
- González-Rangel, A. A. & Navarro, R. E. 2021. Lin-35 beyond its classical roles: Its function in the stress response. *The International journal of developmental biology*, 65, 377-382.
- Gosling, S. D. & John, O. P. 1999. Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science*, 8, 69-75.
- Gosling, S. D., Kwan, V. S. & John, O. P. 2003. A dog's got personality: A cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology*, 85, 1161-9.
- Gouveia, C., Gibbons, E., Dehghani, N., Eapen, J., Guerreiro, R., et al. 2022. Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank. *Scientific reports*, 12, 8404.
- Grandin, T. & Deesing, M. J. 2013. *Genetics and the behavior of domestic animals*, Academic press.
- Gray, A. 2020. Impending dog behaviour crisis following COVID-19 lockdown. *Veterinary Record*, 187.
- Greenwood, T. A., Lazzeroni, L. C., Maihofer, A. X., Swerdlow, N. R., Calkins, M. E., et al. 2019. Genome-wide association of endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia (COGS) study. *JAMA psychiatry*, 76, 1274-1284.
- Guide Dogs. 2020a. *Training with a guide dog* [Online]. Available: <https://www.guidedogs.org.uk/getting-support/guide-dogs/training-with-a-guide-dog#how-are-guide-dogs-trained> [Accessed 22 December 2020].
- Guide Dogs. 2020b. *Born to guide project* [Online]. Available: <https://www.guidedogs.org.uk/about-us/what-we-do/research/born-to-guide-project> [Accessed 18 December 2020].
- Guide Dogs. 2025. *How your money is helping* [Online]. Available: <https://www.guidedogs.org.uk/about-us/how-your-money-is-helping/> [Accessed 08 April 2025].
- Guo, H., Peng, Y., Hu, Z., Li, Y., Xun, G., et al. 2017. Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. *Scientific reports*, 7, 44155.
- Guo, L., Liu, Z., Jia, X., Wang, Q., Ji, J., et al. 2024. Mitochondrial protein tamm41 modulates depressive-like behaviors. *Mol Neurobiol*.
- Gupta, P., Galimberti, M., Liu, Y., Beck, S., Wingo, A., et al. 2024. A genome-wide investigation into the underlying genetic architecture of personality traits and overlap with psychopathology. *Nature Human Behaviour*, 1-15.
- Hall, N. J. & Wynne, C. D. L. 2012. The canid genome: Behavioral geneticists' best friend? *Genes, Brain and Behavior*, 11, 889-902.
- Handegård, K. W., Storengen, L. M., Joergensen, D. & Lingaas, F. 2023. Genomic analysis of firework fear and noise reactivity in standard poodles. *Canine Medicine and Genetics*, 10, 2.

- Hare, E., Essler, J. L., Otto, C. M., Ebbecke, D. & Serpell, J. A. 2024. Development of a modified c-barq for evaluating behavior in working dogs. *Frontiers in Veterinary Science*, 11, 1371630.
- Harper, J. D., Fan, K.-H., Aslam, M. M., Snitz, B. E., DeKosky, S. T., et al. 2022. Genome-wide association study of incident dementia in a community-based sample of older subjects. *Journal of Alzheimer's Disease*, 88, 787-798.
- Harro, J. 2018. Animals, anxiety, and anxiety disorders: How to measure anxiety in rodents and why. *Behavioural brain research*, 352, 81-93.
- Harvey, N. D. 2014. *Development and evaluation of methods for determining personality in juvenile guide dogs*. PhD, Univeristy of Nottingham.
- Harvey, N. D., Craigon, P. J., Blythe, S. A., England, G. C. W. & Asher, L. 2016. Social rearing environment influences dog behavioral development. *Journal of Veterinary Behavior*, 16, 13-21.
- Harvey, N. D., Craigon, P. J., Blythe, S. A., England, G. C. W. & Asher, L. 2017. An evidence-based decision assistance model for predicting training outcome in juvenile guide dogs. *PloS one*, 12, e0174261.
- Harvey, N. D. 2021. A simple guide to inter-rater, intra-rater and test-retest reliability for animal behaviour studies. *OSF Preprints*.
- Haverbeke, A., De Smet, A., Depiereux, E., Giffroy, J.-M. & Diederich, C. 2009. Assessing undesired aggression in military working dogs. *Applied Animal Behaviour Science*, 117, 55-62.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92, 433-443.
- Hayward, J. J., Castelhana, M. G., Oliveira, K. C., Corey, E., Balkman, C., et al. 2016. Complex disease and phenotype mapping in the domestic dog. *Nature communications*, 7, 10460.
- Hayward, J. J., White, M. E., Boyle, M., Shannon, L. M., Casal, M. L., et al. 2019. Imputation of canine genotype array data using 365 whole-genome sequences improves power of genome-wide association studies. *PLoS Genetics*, 15, e1008003.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics*, 28, 476-490.
- He, L., Loika, Y., Park, Y., consortium, G. T. E., Bennett, D. A., et al. 2021a. Exome-wide age-of-onset analysis reveals exonic variants in *ern1* and *sppl2c* associated with alzheimer's disease. *Translational Psychiatry*, 11, 146.
- He, Y., Vinkers, C. H., Houtepen, L. C., De Witte, L. D. & Boks, M. P. 2019. Childhood adversity is associated with increased *kitlg* methylation in healthy individuals but not in bipolar disorder patients. *Frontiers in psychiatry*, 9, 743.
- He, Y., Tiezzi, F., Howard, J. & Maltecca, C. 2021b. Predicting body weight in growing pigs from feeding behavior data using machine learning algorithms. *Computers and Electronics in Agriculture*, 184, 106085.
- Hedges, S. 2015. Advanced approaches to handling dogs in practice. *The Veterinary Nurse*, 6, 308-315.

- Held, S., Mendl, M., Laughlin, K. & Byrne, R. 2002. Cognition studies with pigs: Livestock cognition and its implication for production. *Journal of Animal Science*, 80, E10-E17.
- Helton, W. S. 2010. Does perceived trainability of dog (*canis lupus familiaris*) breeds reflect differences in learning or differences in physical ability? *Behavioural processes*, 83, 315-323.
- Hens, K. 2009. Ethical responsibilities towards dogs: An inquiry into the dog–human relationship. *Journal of Agricultural and Environmental Ethics*, 22, 3-14.
- Hiby, E., Rooney, N. & Bradshaw, J. 2004. Dog training methods: Their use, effectiveness and interaction with behaviour and welfare. *Animal Welfare*, 13, 63-69.
- Hjelm-Björkman, A. K., Kapatkin, A. S. & Rita, H. J. 2011. Reliability and validity of a visual analogue scale used by owners to measure chronic pain attributable to osteoarthritis in their dogs. *American Journal of Veterinary Research*, 72, 601-7.
- Hindley, G., Shadrin, A. A., van der Meer, D., Parker, N., Cheng, W., et al. 2023. Multivariate genetic analysis of personality and cognitive traits reveals abundant pleiotropy. *Nature Human Behaviour*, 7, 1584-1600.
- Hoffman, C. L., Chen, P., Serpell, J. A. & Jacobson, K. C. 2013. Do dog behavioral characteristics predict the quality of the relationship between dogs and their owners? *Human-animal interaction bulletin*.
- Hölter, S. M., Garrett, L., Einicke, J., Sperling, B., Dirscherl, P., et al. 2015. Assessing cognition in mice. *Current protocols in mouse biology*, 5, 331-358.
- Homann, J., Osburg, T., Ohlei, O., Dobricic, V., Deecke, L., et al. 2022. Genome-wide association study of alzheimer's disease brain imaging biomarkers and neuropsychological phenotypes in the european medical information framework for alzheimer's disease multimodal biomarker discovery dataset. *Frontiers in aging neuroscience*, 14, 840651.
- Hoover, K. C. 2010. Smell with inspiration: The evolutionary significance of olfaction. *American Journal of Physical Anthropology*, 143, 63-74.
- Hope, T. M., Price, C. J., Halai, A., Salvi, C., Crinion, J., et al. 2024. Estimating the construct validity of principal components analysis. *arXiv preprint arXiv:2401.12905*.
- Hopkins, W. D., Russell, J. L. & Schaeffer, J. 2014. Chimpanzee intelligence is heritable. *Current Biology*, 24, 1649-1652.
- Hoppe, N., Bininda-Emonds, O. & Gansloßer, U. 2017. Correlates of attention deficit hyperactivity disorder (adhd)-like behavior in domestic dogs: First results from a questionnaire-based study. *Open Journal of Veterinary Medicine*, 2, 95-131.
- Horowitz, A., Hecht, J. & Dedrick, A. 2013. Smelling more or less: Investigating the olfactory experience of the domestic dog. *Learning and motivation*, 44, 207-217.
- Howard, D. M., Adams, M. J., Shirali, M., Clarke, T.-K., Marioni, R. E., et al. 2018. Genome-wide association study of depression phenotypes in uk biobank identifies variants in excitatory synaptic pathways. *Nature communications*, 9, 1470.

- Howell, T. J. & Bennett, P. C. 2011. Puppy power! Using social cognition research tasks to improve socialization practices for domestic dogs (*canis familiaris*). *Journal of Veterinary Behavior*, 6, 195-204.
- Hradecká, L., Bartoš, L., Svobodová, I. & Sales, J. 2015. Heritability of behavioural traits in domestic dogs: A meta-analysis. *Applied Animal Behaviour Science*, 170, 1-13.
- Hsu, Y. & Serpell, J. A. 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association*, 223, 1293-300.
- Huckins, L. M., Logan, D. W. & Sánchez-Andrade, G. 2013. Olfaction and olfactory-mediated behaviour in psychiatric disease models. *Cell and tissue research*, 354, 69-80.
- Humphrey, E. & Warner, L. H. 1934. *Working dogs: An attempt to produce a strain of german shepherds which combines working ability and beauty of conformation, by elliot humphrey and lubien warner; with a foreword by raymond pearl*, Johns Hopkins Press.
- Hunt, R. L., England, G. C. W., Asher, L., Whiteside, H. & Harvey, N. D. 2020. Concurrent and predictive criterion validity of a puppy behaviour questionnaire for predicting training outcome in juvenile guide dogs. *Animals (Basel)*, 10.
- Ignatieva, E. V., Afonnikov, D. A., Saik, O. V., Rogaev, E. I. & Kolchanov, N. A. 2016. A compendium of human genes regulating feeding behavior and body weight, its functional characterization and identification of gwas genes involved in brain-specific ppi network. *BMC Genetics*, 17, 89-116.
- Ikeda, M., Takahashi, A., Kamatani, Y., Momozawa, Y., Saito, T., et al. 2019. Genome-wide association study detected novel susceptibility genes for schizophrenia and shared trans-populations/diseases genetic effect. *Schizophrenia bulletin*, 45, 824-834.
- Ilska, J., Haskell, M. J., Blott, S. C., Sánchez-Molano, E., Polgar, Z., et al. 2017. Genetic characterization of dog personality traits. *Genetics (Austin)*, 206, 1101-1111.
- IWDBA. 2025. *Behavior check list* [Online]. Available: <https://www.iwdba.org/working-dog-resources/list-of-all-bcl-items/> [Accessed 10 April 2025].
- Jagannathan, V., Hitte, C., Kidd, J. M., Masterson, P., Murphy, T. D., et al. 2021. Dog10k_boxer_tasha_1.0: A long-read assembly of the dog reference genome. *Genes*, 12, 847.
- Jakuba, T., Polcová, Z., Fedáková, D., Kottferová, J., Mareková, J., et al. 2013. Differences in evaluation of a dog's temperament by individual members of the same household. *Society & Animals*, 21, 582-589.
- Janssen, F., Bardoutsos, A. & Vidra, N. 2020. Obesity prevalence in the long-term future in 18 european countries and in the USA. *Obesity facts*, 13, 514-527.
- Janutta, V., Hamann, H. & Distl, O. 2006. Complex segregation analysis of canine hip dysplasia in german shepherd dogs. *Journal of Heredity*, 97, 13-20.

- Jarvis, S. 2018. Encouraging responsible pet ownership. *Vet Record*, 182, 389.
- Jendryn, P., Schulz, C., Twele, F., Meller, S., von Köckritz-Blickwede, M., et al. 2020. Scent dog identification of samples from covid-19 patients—a pilot study. *BMC infectious diseases*, 20, 1-7.
- Jenkins, C. A., Consortium, D. B. V. D., Aguirre, G., André, C., Bannasch, D., et al. 2021. Improving the resolution of canine genome-wide association studies using genotype imputation: A study of two breeds. *Animal Genetics*, 52, 703-713.
- Jensen, P. & Andersson, L. 2005. Genomics meets ethology: A new route to understanding domestication, behavior, and sustainability in animal breeding. *Ambio*, 320-324.
- Jensen, P., Buitenhuis, B., Kjaer, J., Zanella, A., Mormède, P., et al. 2008. Genetics and genomics of animal behaviour and welfare—challenges and possibilities. *Applied Animal Behaviour Science*, 113, 383-403.
- Jia, P., Zhao, Z., Hulkan, T., Bush, W. S., Samuels, D. C., et al. 2017. Genome-wide association study of hiv-associated neurocognitive disorder (hand): A charter group study. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 174, 413-426.
- Jolliffe, I. T. & Cadima, J. 2016. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374, 20150202.
- Jones, A. C. & Gosling, S. D. 2005. Temperament and personality in dogs (*canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science*, 95, 1-53.
- Jones, A. C. 2009. *Development and validation of a dog personality questionnaire*. PhD, University of Texas.
- Jordan, M. I. & Mitchell, T. M. 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255-260.
- Joshi, A. 2005. Behaviour genetics in the post-genomics era: From genes to behaviour and vice versa. *Current Science*, 1128-1135.
- Jukan, A., Masip-Bruin, X. & Amla, N. 2017. Smart computing and sensing technologies for animal welfare: A systematic review. *ACM Computing Surveys (CSUR)*, 50, 1-27.
- Kagan, J. 1997. Temperament and the reactions to unfamiliarity. *Child development*, 68, 139-143.
- Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Kang, M., Ang, T. F. A., Devine, S. A., Sherva, R., Mukherjee, S., et al. 2023. A genome-wide search for pleiotropy in more than 100,000 harmonized longitudinal cognitive domain scores. *Molecular neurodegeneration*, 18, 40.
- Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., et al. 2019. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, 51, 245-257.
- Karlsson Linnér, R., Mallard, T. T., Barr, P. B., Sanchez-Roige, S., Madole, J. W., et al. 2021. Multivariate analysis of 1.5 million

- people identifies genetic associations with traits related to self-regulation and addiction. *Nat Neurosci*, 24, 1367-1376.
- Kato, H., Kushima, I., Mori, D., Yoshimi, A., Aleksic, B., et al. 2020. Rare genetic variants in the gene encoding histone lysine demethylase 4c (kdm4c) and their contributions to susceptibility to schizophrenia and autism spectrum disorder. *Translational Psychiatry*, 10, 421.
- Kaur, S., Ali, A., Ahmad, U., Siahbalaei, Y., Pandey, A. K., et al. 2019. Role of single nucleotide polymorphisms (snps) in common migraine. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 55, 47.
- Kavlak, A. T., Pastell, M. & Uimari, P. 2023. Disease detection in pigs based on feeding behaviour traits using machine learning. *biosystems engineering*, 226, 132-143.
- Ke, X., Kennedy, L., Short, A., Seppälä, E., Barnes, A., et al. 2011. Assessment of the functionality of genome-wide canine snp arrays and implications for canine disease association studies. *Animal Genetics*, 42, 181-190.
- Keijser, S., Fieten, H., Vos-Loohuis, M., Piek, C., Anderson, H., et al. 2018a. Heterozygosity testing and multiplex DNA panel screening as a potential tool to monitor health and inbreeding in a small, closed dog population. *Canine Genetics and Epidemiology*, 5, 1-9.
- Keijser, S. F. A., Fieten, H., Vos-Loohuis, M., Piek, C. J., Anderson, H., et al. 2018b. Heterozygosity testing and multiplex DNA panel screening as a potential tool to monitor health and inbreeding in a small, closed dog population. *Canine Genetics and Epidemiology*, 5, 12.
- Kernberg, O. F. 2016. What is personality? *Journal of personality disorders*, 30, 145-156.
- Kessler, R. C. & Greenberg, P. E. 2002. The economic burden of anxiety and stress disorders. *Neuropsychopharmacology: The fifth generation of progress*, 67, 982-92.
- Kim, J. & Moon, N. 2022. Dog behavior recognition based on multimodal data from a camera and wearable device. *Applied sciences*, 12, 3199.
- Kim, S. E., Kim, H.-N., Yun, Y.-J., Heo, S. G., Cho, J., et al. 2017. Meta-analysis of genome-wide snp-and pathway-based associations for facets of neuroticism. *Journal of Human Genetics*, 62, 903-909.
- Kim, S. Y., Chung, H. S., Sun, W. & Kim, H. 2007. Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. *Neuroscience*, 147, 996-1021.
- Kimbrel, N. A., Ashley-Koch, A. E., Qin, X. J., Lindquist, J. H., Garrett, M. E., et al. 2023. Identification of novel, replicable genetic risk loci for suicidal thoughts and behaviors among us military veterans. *JAMA psychiatry*, 80, 135-145.
- King, T., Hemsworth, P. H. & Coleman, G. J. 2003. Fear of novel and startling stimuli in domestic dogs. *Applied Animal Behaviour Science*, 82, 45-64.
- Kirkpatrick, B., Ge, S. & Wang, L. 2018. Efficient computation of the kinship coefficients. *Bioinformatics*, 35, 1002-1008.

- Kleanthous, N., Hussain, A. J., Khan, W., Sneddon, J., Al-Shamma'a, A., et al. 2022. A survey of machine learning approaches in animal behaviour. *Neurocomputing*, 491, 442-463.
- Kobayashi, N., Arata, S., Hattori, A., Kohara, Y., Kiyokawa, Y., et al. 2013. Association of puppies' behavioral reaction at five months of age assessed by questionnaire with their later 'distraction' at 15 months of age, an important behavioral trait for guide dog qualification. *Journal of Veterinary Medical Science*, 75, 63-67.
- Kohl, M., Foulon, C. & Guelfi, J. 2004. Hyperactivity and anorexia nervosa: Behavioural and biological perspective. *L'encephale*, 30, 492-499.
- Kokocińska-Kusiak, A., Woszczyło, M., Zybala, M., Maciocha, J., Barłowska, K., et al. 2021. Canine olfaction: Physiology, behavior, and possibilities for practical applications. *Animals*, 11, 2463.
- Kotov, R., Gamez, W., Schmidt, F. & Watson, D. 2010. Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological bulletin*, 136, 768.
- Krishna, K. K., Hertel, N. & Redies, C. 2011. Cadherin expression in the somatosensory cortex: Evidence for a combinatorial molecular code at the single-cell level. *Neuroscience*, 175, 37-48.
- Krosnick, J. A. 2018. *Questionnaire design*, Palgrave Macmillan Cham.
- Kubinyi, E., Turcsán, B. & Miklósi, Á. 2009. Dog and owner demographic characteristics and dog personality trait associations. *Behavioural processes*, 81, 392-401.
- Kulminski, A. M., Loiko, E., Loika, Y. & Culminkaya, I. 2022. Pleiotropic predisposition to alzheimer's disease and educational attainment: Insights from the summary statistics analysis. *Geroscience*, 1-16.
- Kumar, V., Bhat, Z. A. & Kumar, D. 2013. Animal models of anxiety: A comprehensive review. *Journal of pharmacological and toxicological methods*, 68, 175-183.
- Kumpulainen, P., Cardó, A. V., Somppi, S., Törnqvist, H., Väättäjä, H., et al. 2021. Dog behaviour classification with movement sensors placed on the harness and the collar. *Applied Animal Behaviour Science*, 241, 105393.
- Kuo, P.-H., Chuang, L.-C., Su, M.-H., Chen, C.-H., Chen, C.-H., et al. 2015. Genome-wide association study for autism spectrum disorder in taiwanese han population. *PloS one*, 10, e0138695.
- Kurvers, R. H., Prins, H. H., van Wieren, S. E., van Oers, K., Nolet, B. A., et al. 2010. The effect of personality on social foraging: Shy barnacle geese scrounge more. *Proceedings of the Royal Society B: Biological Sciences*, 277, 601-8.
- Kushima, I., Aleksic, B., Ito, Y., Nakamura, Y., Nakamura, K., et al. 2010. Association study of ubiquitin-specific peptidase 46 (usp46) with bipolar disorder and schizophrenia in a japanese population. *Journal of Human Genetics*, 55, 133-136.
- Kweon, K., Shin, E.-S., Park, K. J., Lee, J.-K., Joo, Y., et al. 2018. Genome-wide analysis reveals four novel loci for attention-deficit hyperactivity disorder in korean youths. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, 29, 62.

- Lampe, J. F., Ruchti, S., Burman, O., Würbel, H. & Melotti, L. 2019. Play like me: Similarity in playfulness promotes social play. *PloS one*, 14, e0224282.
- Lanfredi, M., Macis, A., Ferrari, C., Meloni, S., Pedrini, L., et al. 2021. Maladaptive behaviours in adolescence and their associations with personality traits, emotion dysregulation and other clinical features in a sample of italian students: A cross-sectional study. *Borderline personality disorder and emotion dysregulation*, 8, 14.
- Lavrijsen, I. C. M., Leegwater, P. A. J., Wangdee, C., van Steenbeek, F. G., Schwencke, M., et al. 2014. Genome-wide survey indicates involvement of loci on canine chromosomes 7 and 31 in patellar luxation in flat-coated retrievers. *BMC Genetics*, 15, 64.
- Lazarowski, L., Strassberg, L. R., Waggoner, L. P. & Katz, J. S. 2019. Persistence and human-directed behavior in detection dogs: Ontogenetic development and relationships to working dog success. *Applied Animal Behaviour Science*, 220, 104860.
- Lazarowski, L., Krichbaum, S., Waggoner, L. P. & Katz, J. S. 2020. The development of problem-solving abilities in a population of candidate detection dogs (canis familiaris). *Animal Cognition*, 23, 755-768.
- Lazarowski, L., Rogers, B., Krichbaum, S., Haney, P., Smith, J. G., et al. 2021. Validation of a behavior test for predicting puppies' suitability as detection dogs. *Animals*, 11, 993.
- Le-Niculescu, H., Roseberry, K., Gill, S. S., Levey, D. F., Phalen, P. L., et al. 2021. Precision medicine for mood disorders: Objective assessment, risk prediction, pharmacogenomics, and repurposed drugs. *Mol Psychiatry*, 26, 2776-2804.
- Leblond, C. S., Cliquet, F., Carton, C., Huguet, G., Mathieu, A., et al. 2019. Both rare and common genetic variants contribute to autism in the faroe islands. *NPJ genomic medicine*, 4, 1.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., et al. 2018. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50, 1112-1121.
- Leighton, E. A. 2009. Secrets for producing high-quality working dogs. *Journal of Veterinary Behavior*, 4, 212-215.
- Levey, D. F., Stein, M. B., Wendt, F. R., Pathak, G. A., Zhou, H., et al. 2021. Bi-ancestral depression gwas in the million veteran program and meta-analysis in > 1.2 million individuals highlight new therapeutic directions. *Nat Neurosci*, 24, 954-963.
- Levitis, D. A., Lidicker Jr, W. Z. & Freund, G. 2009. Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour*, 78, 103-110.
- Lewis, T. W., Blott, S. C. & Woolliams, J. A. 2010. Genetic evaluation of hip score in uk labrador retrievers. *PloS one*, 5, e12797-e12797.
- Lewis, T. W., Blott, S. C. & Woolliams, J. A. 2013. Comparative analyses of genetic trends and prospects for selection against hip and elbow dysplasia in 15 uk dog breeds. *BMC Genetics*, 14, 16.
- Lewis, T. W. 2015. The genetics of complex traits – applying theory to selection on behaviour. Abstracts from the canine behaviour and

- genetics meeting, london, june 2015. *Journal of Veterinary Behavior*, 10, 438-448.
- Lewis, T. W. 2016. Improving the efficacy of selection on complex traits—a brief review of quantitative genetic methods and considerations for measurement of phenotypes. *Journal of Veterinary Behavior*, 16, 76-80.
- Lewis, T. W. 2019. The impact of incorrectly recorded parentage on inferred genotypes over multiple generations. *Canine Genetics and Epidemiology*, 6, 6.
- Ley, J., Bennett, P. & Coleman, G. 2008. Personality dimensions that emerge in companion canines. *Applied Animal Behaviour Science*, 110, 305-317.
- Ley, J. M., Bennett, P. C. & Coleman, G. J. 2009. A refinement and validation of the monash canine personality questionnaire (mcpq). *Applied Animal Behaviour Science*, 116, 220-227.
- Li, Z., Chen, J., Yu, H., He, L., Xu, Y., et al. 2017. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature Genetics*, 49, 1576-1583.
- Liinamo, A.-E., van den Berg, L., Leegwater, P. A. J., Schilder, M. B. H., van Arendonk, J. A. M., et al. 2007. Genetic variation in aggression-related traits in golden retriever dogs. *Applied Animal Behaviour Science*, 104, 95-106.
- Lim, J. & Dinges, D. F. 2010. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological bulletin*, 136, 375.
- Lin, E., Kuo, P.-H., Liu, Y.-L., Yu, Y. W.-Y., Yang, A. C., et al. 2018. A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. *Frontiers in psychiatry*, 9, 290.
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438, 803-819.
- Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., et al. 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, 51, 237-244.
- Lush, J. L. 1949. Heritability of quantitative characters in farm animals. *Hereditas*, 35, 356-375.
- Lutz, M. W., Casanova, R., Saldana, S., Kuchibhatla, M., Plassman, B. L., et al. 2019. Analysis of pleiotropic genetic effects on cognitive impairment, systemic inflammation, and plasma lipids in the health and retirement study. *Neurobiology of aging*, 80, 173-186.
- Ma, P., Huang, J., Gong, W., Li, X., Gao, H., et al. 2018. The impact of genomic relatedness between populations on the genomic estimated breeding values. *Journal of animal science and biotechnology*, 9, 1-6.
- Mackenzie, S. A., Oltenacu, E. A. B. & Leighton, E. 1985. Heritability estimate for temperament scores in german shepherd dogs and its genetic correlation with hip dysplasia. *Behavior Genetics*, 15, 475-482.

- Mackenzie, S. A., Oltenacu, E. A. B. & Houpt, K. A. 1986. Canine behavioral genetics — a review. *Applied Animal Behaviour Science*, 15, 365-393.
- MacLean, E. L. & Bray, E. E. 2019. Ontogeny and heritability of cognitive and behavioral traits linked to success as a military working dog. DUKE UNIVERSITY Durham United States.
- MacLean, E. L., Snyder-Mackler, N., VonHoldt, B. M. & Serpell, J. A. 2019. Highly heritable and functionally relevant breed differences in dog behaviour. *Proceedings of the Royal Society B*, 286, 20190716.
- Mai, D. L., Howell, T., Benton, P. & Bennett, P. C. 2021. Socialisation, training, and help-seeking—specific puppy raising practices that predict desirable behaviours in trainee assistance dog puppies. *Applied Animal Behaviour Science*, 236, 105259.
- Malhotra, N. K. 2006. *Questionnaire design*, SAGE Publications, Inc.
- Mallard, B. A., Wilkie, B. N., Kennedy, B. W. & Quinton, M. 1992. Use of estimated breeding values in a selection index to breed yorkshire pigs for high and low immune and innate resistance factors. *Animal Biotechnology*, 3, 257-280.
- Malmkvist, J. & Hansen, S. 2001. The welfare of farmed mink (*Mustela vison*) in relation to behavioural selection: A review. *Animal Welfare*, 10, 41-52.
- Marcato, M., Kenny, J., O'Riordan, R., O'Mahony, C., O'Flynn, B., et al. 2022. Assistance dog selection and performance assessment methods using behavioural and physiological tools and devices. *Applied Animal Behaviour Science*, 254.
- Marshall-Pescini, S., Valsecchi, P., Petak, I., Accorsi, P. A. & Previde, E. P. 2008. Does training make you smarter? The effects of training on dogs' performance (*Canis familiaris*) in a problem solving task. *Behavioural processes*, 78, 449-454.
- Marth, G. T., Korf, I., Yandell, M. D., Yeh, R. T., Gu, Z., et al. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics*, 23, 452-456.
- Matsumoto, Y., Konno, A., Ishihara, G. & Inoue-Murayama, M. 2023. Genetic dissection of behavioral traits related to successful training of drug detection dogs. *Scientific reports*, 13, 7326.
- McCrae, R. R. & John, O. P. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175-215.
- McCrae, R. R., Costa, P. T., Jr., Ostendorf, F., Angleitner, A., Hrebícková, M., et al. 2000. Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78, 173-86.
- McGarrity, M. E., Sinn, D. L., Thomas, S. G., Marti, C. N. & Gosling, S. D. 2016. Comparing the predictive validity of behavioral codings and behavioral ratings in a working-dog breeding program. *Applied Animal Behaviour Science*, 179, 82-94.
- McHugh, M. L. 2012. Interrater reliability: The kappa statistic. *Biochemia medica*, 22, 276-282.

- McIver, S., Hall, S. & Mills, D. S. 2020. The impact of owning a guide dog on owners' quality of life: A longitudinal study. *Anthrozoös*, 33, 103-117.
- Meagher, R. K. 2009. Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science*, 119, 1-14.
- Mehrkam, L. R. & Wynne, C. D. L. 2014. Behavioral differences among breeds of domestic dogs (*canis lupus familiaris*): Current status of the science. *Applied Animal Behaviour Science*, 155, 12-27.
- Mehta, P. H. & Gosling, S. D. 2008. Bridging human and animal research: A comparative approach to studies of personality and health. *Brain, Behavior, and Immunity*, 22, 651-661.
- Mellersh, C. 2008. Give a dog a genome. *The Veterinary Journal*, 178, 46-52.
- Mendl, M., Burman, O. H. & Paul, E. S. 2010. An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, 277, 2895-2904.
- Meneses, T., Robinson, J., Rose, J., Vernick, J. & Overall, K. L. 2021. Review of epidemiological, pathological, genetic, and epigenetic factors that may contribute to the development of separation anxiety in dogs. *Journal of the American Veterinary Medical Association*, 259, 1118-1129.
- Meng, X., Navoly, G., Giannakopoulou, O., Levey, D. F., Koller, D., et al. 2024. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nature Genetics*, 56, 222-233.
- Menor-Campos, D. J. 2024. Ethical concerns about fashionable dog breeding. *Animals*, 14, 756.
- Menuge, F., Marcet-Rius, M., Jochem, M., François, O., Assali, C., et al. 2021. Early evaluation of fearfulness in future guide dogs for blind people. *Animals*, 11, 412.
- Meuwissen, T. H. E. & Luo, Z. 1992. Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, 24, 305.
- Meyers-Wallen, V. 2003. Ethics and genetic selection in purebred dogs. *Reproduction in domestic animals*, 38, 73-76.
- Mez, J., Chung, J., Jun, G., Kriegel, J., Bourlas, A. P., et al. 2017. Two novel loci, *cobl* and *slc10a2*, for alzheimer's disease in african americans. *Alzheimer's and Dementia*, 13, 119-129.
- Micioni Di Bonaventura, E., Botticelli, L., Tomassoni, D., Tayebati, S. K., Micioni Di Bonaventura, M. V., et al. 2020. The melanocortin system behind the dysfunctional eating behaviors. *Nutrients*, 12, 3502.
- Miletto Petrazzini, M. E., Mantese, F. & Prato-Previde, E. 2020. Food quantity discrimination in puppies (*canis lupus familiaris*). *Animal Cognition*, 23, 703-710.
- Mirkó, E., Kubinyi, E., Gácsi, M. & Miklósi, Á. 2012. Preliminary analysis of an adjective-based dog personality questionnaire developed to measure some aspects of personality in the domestic dog (*canis familiaris*). *Applied Animal Behaviour Science*, 138, 88-98.

- Mizukoshi, M., Kondo, M. & Nakamura, T. 2008. Evaluation of the potential suitability of guide dog candidates by continuous observation during training. *Journal of Veterinary Behavior*, 3, 193-198.
- Montibeller, L. & de Belleruche, J. 2018. Amyotrophic lateral sclerosis (als) and alzheimer's disease (ad) are characterised by differential activation of er stress pathways: Focus on upr target genes. *Cell Stress Chaperones*, 23, 897-912.
- Morel, E., Malineau, L., Venet, C., Gaillard, V. & Péron, F. 2024. Prioritization of appearance over health and temperament is detrimental to the welfare of purebred dogs and cats. *Animals*, 14, 1003.
- Mornement, K., Toukhsati, S., Coleman, G. & Bennett, P. Reliability, validity and feasibility of existing tests of canine behaviour. AIAM Annual Conference on Urban Animal Management, Proceedings, 2009. 11-18.
- Morrill, K., Hekman, J., Li, X., McClure, J., Logan, B., et al. 2022. Ancestry-inclusive dog genomics challenges popular breed stereotypes. *Science*, 376, eabk0639.
- Moser, G., Tier, B., Crump, R. E., Khatkar, M. S. & Raadsma, H. W. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide snp markers. *Genetics Selection Evolution*, 41, 56.
- Moxon, R. 2018. Outcome of all dogs born 2015. Guide Dogs.
- Moxon, R. & Fellows, H. 2022. Outcome of all dogs born in 2019. Guide Dogs Internal Report.
- Moxon, R., Freeman, S. L., Payne, R., Corr, S. & England, G. C. 2023. Effect of neutering timing in relation to puberty on female dog behaviour—a scoping review. *Veterinary Record*, 193, no-no.
- Moxon, R. & Fellows, H. 2024. Outcome of all dogs born in 2020. Guide Dogs for the Blind.
- Mukaka, M. 2012. Statistics corner: A guide to appropriate use of correlation in medical research. *Malawi Medical Journal*, 24, 69-71.
- Munch, K. L., Wapstra, E., Thomas, S., Fisher, M. & Sinn, D. L. 2019. What are we measuring? Novices agree amongst themselves (but not always with experts) in their assessment of dog behaviour. *Ethology*, 125, 203-211.
- Murphy, J. A. 1995. Assessment of the temperament of potential guide dogs. *Anthrozoös*, 8, 224-228.
- Murphy, J. A. 1998. Describing categories of temperament in potential guide dogs for the blind. *Applied Animal Behaviour Science*, 58, 163-178.
- Naganuma, F., Nakamura, T., Yoshikawa, T., Iida, T., Miura, Y., et al. 2017. Histamine n-methyltransferase regulates aggression and the sleep-wake cycle. *Science Reports*, 7, 15899.
- Nagasawa, M., Kanbayashi, S., Mogi, K., Serpell, J. A. & Kikusui, T. 2016. Comparison of behavioral characteristics of dogs in the united states and japan. *Journal of Veterinary Medical Science*, 78, 231-238.

- Naj, A. C., Beecham, G. W., Martin, E. R., Gallins, P. J., Powell, E. H., et al. 2010. Dementia revealed: Novel chromosome 6 locus for late-onset alzheimer disease provides genetic evidence for folate-pathway abnormalities. *PLoS Genetics*, 6, e1001130.
- Naskar, T., Faruq, M., Banerjee, P., Khan, M., Midha, R., et al. 2018. Ancestral variations of the pcdhg gene cluster predispose to dyslexia in a multiplex family. *EBioMedicine*, 28, 168-179.
- Navarro, M. 2017. The role of the melanocortin system in drug and alcohol abuse. *International review of neurobiology*, 136, 121-150.
- NCBI. 2020. *Genome assembly ros_cfam_1.0* [Online]. Available: https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_014441545.1/ [Accessed 03/08/2024 2024].
- Ng, P. C. & Kirkness, E. F. 2010. Whole genome sequencing. *Genetic variation*, 215-226.
- Nose, I. & Kakinuma, M. 2019. Japanese version of the dog personality scale: Reliability and validity based on owner and dog instructor evaluation and behavioral observation. *Applied Animal Behaviour Science*, 214, 72-78.
- Nuttall, T. 2013. The genomics revolution: Will canine atopic dermatitis be predictable and preventable? *Vet Dermatol*, 24, 10-8.e3-4.
- O'Neill, D. G. & Packer, R. M. 2016. The first canine behavior and genetics conference: Summary and recommendations for future directions in canine behavioral science. *Journal of Veterinary Behavior*, 16, 6-12.
- Ogata, N. 2016. Separation anxiety in dogs: What progress has been made in our understanding of the most common behavioral problems in dogs? *Journal of Veterinary Behavior*, 16, 28-35.
- Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., et al. 2022. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*, 54, 437-449.
- Ollagnier, C., Kasper, C., Wallenbeck, A., Keeling, L., Bee, G., et al. 2023. Machine learning algorithms can predict tail biting outbreaks in pigs using feeding behaviour records. *PloS one*, 18, e0252002.
- Olsen, A. M., Evans, R. B. & Duerr, F. M. 2016. Evaluation of accelerometer inter-device variability and collar placement in dogs. *Veterinary Evidence*, 1.
- Oroszi, G., Enoch, M. A., Chun, J., Virkkunen, M. & Goldman, D. 2005. Thr105ile, a functional polymorphism of histamine n-methyltransferase, is associated with alcoholism in two independent populations. *Alcoholism: Clinical and Experimental Research*, 29, 303-309.
- Osborne, J. W. & Costello, A. B. 2019. Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, 9, 11.
- Ostrander, E. A. & Wayne, R. K. 2005. The canine genome. *Genome research*, 15, 1706-1716.
- Ostrander, E. A. & Williams, F. 2019. Canine genomics and genetics. *Reference Module in Life Sciences*.
- Otsuka, R., Naganuma, F., Nakamura, T., Miwa, H., Nakayama-Naono, R., et al. 2022. Contribution of astrocytic histamine n-

- methytransferase to histamine clearance and brain function in mice. *Neuropharmacology*, 212, 109065.
- Overall, K. L., Tiira, K., Broach, D. & Bryant, D. 2014. Genetics and behavior: A guide for practitioners. *Veterinary Clinics of North America-Small Animal Practice*, 44, 483-+.
- Overall, K. L., Dunham, A. E. & Juarbe-Diaz, S. V. 2016. Phenotypic determination of noise reactivity in 3 breeds of working dogs: A cautionary tale of age, breed, behavioral assessment, and genetics. *Journal of Veterinary Behavior*, 16, 113-125.
- Paroz, C., Gebhardt-Henrich, S. G. & Steiger, A. 2008. Reliability and validity of behaviour tests in hovawart dogs. *Applied Animal Behaviour Science*, 115, 67-81.
- PDSA. 2024. *Paw report* [Online]. Available: https://www.pdsa.org.uk/media/14944/pdsa_paw-report-2024.pdf [Accessed 2024].
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al. 2011. Scikit-learn: Machine learning in python. *Journal of machine Learning research*, 12, 2825-2830.
- Penke, L., Denissen, J. J. A. & Miller, G. F. 2007. The evolutionary genetics of personality. *European Journal of Personality*, 21, 549-587.
- Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., et al. 2022. Slep: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19, 486-495.
- Perroud, N., Uher, R., Ng, M., Guipponi, M., Hauser, J., et al. 2012. Genome-wide association study of increasing suicidal ideation during antidepressant treatment in the gendep project. *The pharmacogenomics journal*, 12, 68-77.
- Persson, M. E., Roth, L. S. V., Johnsson, M., Wright, D. & Jensen, P. 2015. Human-directed social behaviour in dogs shows significant heritability. *Genes Brain and Behavior*, 14, 337-344.
- Persson, M. E., Wright, D., Roth, L. S. V., Batakis, P. & Jensen, P. 2016. Genomic regions associated with interspecies communication in dogs contain genes related to human social disorders. *Scientific reports*, 6, 33439.
- Pfaffenberger, C. J. & Scott, J. P. 1959. The relationship between delayed socialization and trainability in guide dogs. *The Journal of Genetic Psychology*, 95, 145-155.
- Pinelli, C., Scandurra, A., Di Lucrezia, A., Aria, M., Semin, G. R., et al. 2023. Puppies in the problem-solving paradigm: Quick males and social females. *Animal Cognition*, 26, 791-797.
- Plant, H. 2016. *An evaluation of the predictive value of temperament tests carried out on assistance hearing dogs*. MSc by Research, University of Lincoln.
- Plassais, J., Kim, J., Davis, B. W., Karyadi, D. M., Hogan, A. N., et al. 2019. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nature communications*, 10, 1-14.
- Plomin, R. 1990. The role of inheritance in behavior. *Science*, 248, 183-188.

- Plotnine. n.d. *Plotnine: Grammar of graphics for python* [Online]. Available: <https://plotnine.org/reference/> [Accessed 19th May 2024].
- Price, E. O. 1998. Behavioral genetics and the process of animal domestication. *Genetics and the behavior of domestic animals.*, 31-66.
- Price, M. P., Gong, H., Parsons, M. G., Kundert, J. R., Reznikov, L. R., et al. 2014. Localization and behaviors in null mice suggest that *asic1* and *asic2* modulate responses to aversive stimuli. *Genes, Brain and Behavior*, 13, 179-94.
- Proschowsky, H. F., Arendt, M. L., Bonnett, B. N., Bruun, C. S., Czycholl, I., et al. 2025. A new future for dog breeding. *Animal Welfare*, 34, e1.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., et al. 2007. Plink: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559-575.
- Putnam, S. P. 2011. Stability and instability of childhood traits: Implications for personality development of animals. *Developmental Psychobiology*, 53, 510-520.
- Python. 2018. *Python language reference* [Online]. Python Software Foundation. Available: <https://docs.python.org/3.9/reference/index.html> [Accessed 19th May 2024].
- Radner, W., Diendorfer, G., Kainrath, B. & Kollmitzer, C. 2017. The accuracy of reading speed measurement by stopwatch versus measurement with an automated computer program (rad-rd©). *Acta ophthalmologica*, 95, 211-216.
- Raffan, E., Dennis, R. J., O'Donovan, C. J., Becker, J. M., Scott, R. A., et al. 2016. A deletion in the canine *pomc* gene is associated with weight and appetite in obesity-prone labrador retriever dogs. *Cell metabolism*, 23, 893-900.
- Ralph, R. 2015. Effect of inter and intraobserver variation and observer personality type on a puppy behavioural assessment. University of Nottingham.
- Randall, L. V., Kim, D.-H., Abdelrazig, S. M., Bollard, N. J., Hemingway-Arnold, H., et al. 2023. Predicting lameness in dairy cattle using untargeted liquid chromatography–mass spectrometry-based metabolomics and machine learning. *Journal of Dairy Science*, 106, 7033-7042.
- Rayment, D. J., Peters, R. A., Marston, L. C. & Groef, B. D. 2016. Investigating canine personality structure using owner questionnaires measuring pet dog behaviour and personality. *Applied Animal Behaviour Science*, 180, 100-106.
- Real, R., Martinez-Carrasco, A., Reynolds, R. H., Lawton, M. A., Tan, M. M., et al. 2023. Association between the *lrp1b* and *apoe* loci and the development of parkinson's disease dementia. *Brain*, 146, 1873-1887.
- Réale, D., Martin, J., Coltman, D., Poissant, J. & Festa-Bianchet, M. 2009. Male personality, life-history strategies and reproductive success in

- a promiscuous mammal. *Journal of evolutionary biology*, 22, 1599-1607.
- Reich, D. E., Gabriel, S. B. & Altshuler, D. 2003. Quality and completeness of snp databases. *Nature Genetics*, 33, 457-458.
- Reisner, I. R. 2003. Differential diagnosis and management of human-directed aggression in dogs. *Veterinary Clinics: Small Animal Practice*, 33, 303-320.
- Reitz, N., Kuhne, F. & Struwe, R. 2009. Do different breeding lines show unequal problem-solving strategies at an early age of life?—a comparison of two labrador retriever breeding lines at week 8 of life. *Journal of Veterinary Behavior: Clinical Applications and Research*, 2, 52.
- Rensing, S., Pasman, E. & Reinhardt, F. 2009. Best use of conventional ebv of bull dams and combination with direct genomic values. *Interbull Bulletin*, 123.
- Riemer, S., Müller, C., Virányi, Z., Huber, L. & Range, F. 2013. Choice of conflict resolution strategy is linked to sociability in dog puppies. *Applied Animal Behaviour Science*, 149, 36-44.
- Rigterink, A. & Houpt, K. 2014. Genetics of canine behavior: A review. *World Journal of Medical Genetics*, 4, 46-57.
- Riser, M., Russenberger, J., Moeser, C., Leighton, E. & Frageomeni, B. O. Using single-step genomic blup to compute genomic enhanced breeding values for behavior traits in working dogs. International Working Dog Breeding Conference, 12 – 14 October 2021 2021 Virtual.
- Roberts, S. C., Havlíček, J. & Schaal, B. 2020. Human olfactory communication: Current challenges and future prospects. *Philosophical Transactions of the Royal Society B*, 375, 20190258.
- Rodenburg, T. B. & Turner, S. P. 2012. The role of breeding and genetics in the welfare of farm animals. *Animal Frontiers*, 2, 16-21.
- Rodriguez, K. E., Greer, J., Yacilla, J. K., Beck, A. M. & O’Haire, M. E. 2020. The effects of assistance dogs on psychosocial health and wellbeing: A systematic literature review. *PloS one*, 15, e0243302.
- Rooney, N. & Sargan, D. 2009. Pedigree dog breeding in the uk: A major welfare concern? RSPCA.
- Rooney, N. J., Clark, C. C. & Casey, R. A. 2016. Minimizing fear and anxiety in working dogs: A review. *Journal of Veterinary Behavior*, 16, 53-64.
- Rossi, J. 2013. To breed or not to breed? Selective breeding and the ethics of animal disease.
- Rothbart, M. K., Ahadi, S. A. & Evans, D. E. 2000. Temperament and personality: Origins and outcomes. *Journal of Personality and Social Psychology*, 78, 122-35.
- Ruefenacht, S., Gebhardt-Henrich, S., Miyake, T. & Gaillard, C. 2002. A behaviour test on german shepherd dogs: Heritability of seven different traits. *Applied Animal Behaviour Science*, 79, 113-132.
- Russenberger, J. & Clothier, S. 2008. Novel temperament and relationship assessment tools. *Journal of Veterinary Behavior: Clinical Applications and Research*, 4, 181-182.

- Russenberger, J. 2019. International working dog registry workshops i & ii. *International Working Dog Conference*. Stockhol, Sweden.
- Sachs-Ericsson, N., Hansen, N. K. & Fitzgerald, S. 2002. Benefits of assistance dogs: A review. *Rehabilitation psychology*, 47, 251.
- Saetre, P., Strandberg, E., Sundgren, P. E., Pettersson, U., Jazin, E., et al. 2006. The genetic contribution to canine personality. *Genes, Brain and Behavior*, 5, 240-248.
- Salena, M. G., Turko, A. J., Singh, A., Pathak, A., Hughes, E., et al. 2021. Understanding fish cognition: A review and appraisal of current practices. *Animal Cognition*, 24, 395-406.
- Salmon, A., Driscoll, C., Paterson, M. B., Harpur, P. & Pachana, N. A. 2022. Issues regarding the welfare of assistance dogs. *Animals*, 12, 3250.
- Salonen, M., Sulkama, S., Mikkola, S., Puurunen, J., Hakanen, E., et al. 2020. Prevalence, comorbidity, and breed differences in canine anxiety in 13,700 finnish pet dogs. *Scientific reports*, 10, 2962.
- Sanchez-Roige, S., Gray, J. C., MacKillop, J., Chen, C. H. & Palmer, A. A. 2018. The genetics of human personality. *Genes, Brain and Behavior*, 17, e12439.
- Sandberg, J. V., Hansson, C., Göteson, A., Joas, E., Jakobsson, J., et al. 2022. Proteins associated with future suicide attempts in bipolar disorder: A large-scale biomarker discovery study. *Mol Psychiatry*, 27, 3857-3863.
- Sarviaho, R., Hakosalo, O., Tiira, K., Sulkama, S., Salmela, E., et al. 2019. Two novel genomic regions associated with fearfulness in dogs overlap human neuropsychiatric loci. *Translational Psychiatry*, 9.
- Sarviaho, R., Hakosalo, O., Tiira, K., Sulkama, S., Niskanen, J. E., et al. 2020. A novel genomic region on chromosome 11 associated with fearfulness in dogs. *Translational Psychiatry*, 10, 169.
- Saunders, G. R., Wang, X., Chen, F., Jang, S.-K., Liu, M., et al. 2022. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*, 612, 720-724.
- Scheurink, A. J., Boersma, G. J., Nergårdh, R. & Södersten, P. 2010. Neurobiology of hyperactivity and reward: Agreeable restlessness in anorexia nervosa. *Physiology and behavior*, 100, 490-495.
- Schiefelbein, K. M. 2012. *Estimation of genetic parameters for behavioral assessment scores in labrador retrievers, german shepherd dogs, and golden retrievers*. Master of Science, Kansas State University.
- Schiele, M. A., Bandelow, B., Baldwin, D. S., Pini, S. & Domschke, K. 2020. A neurobiological framework of separation anxiety and related phenotypes. *European neuropsychopharmacology*, 33, 45-57.
- Schneider, L. A., Delfabbro, P. H. & Burns, N. R. 2013. Temperament and lateralization in the domestic dog (canis familiaris). *Journal of Veterinary Behavior*, 8, 124-134.
- Schork, I. G., Manzo, I. A., Oliveira, M. R. B. d., Costa, F. V., Young, R. J., et al. 2023. Testing the accuracy of wearable technology to assess sleep behaviour in domestic dogs: A prospective tool for animal welfare assessment in kennels. *Animals*, 13, 1467.

- Schrauwen, I., Rajendran, Y., Acharya, A., Öhman, S., Arvio, M., et al. 2024. Optical genome mapping unveils hidden structural variants in neurodevelopmental disorders. *Scientific reports*, 14, 11239.
- Schreiber, J. B. 2021. Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, 17, 1004-1011.
- Segurson, S. A., Serpell, J. A. & Hart, B. L. 2005. Evaluation of a behavioral assessment questionnaire for use in the characterization of behavioral problems of dogs relinquished to animal shelters. *Journal of the American Veterinary Medical Association*, 227, 1755-61.
- Serpell, J. A. 1995. *The domestic dog: Its evolution, behaviour and interactions with people*, Cambridge, Cambridge University Press.
- Serpell, J. A. & Hsu, Y. 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science*, 72, 347-364.
- Serpell, J. A. & Hsu, Y. 2005. Effects of breed, sex, and neuter status on trainability in dogs. *Anthrozoös*, 18, 196-207.
- Sforzini, E., Michelazzi, M., Spada, E., Ricci, C., Carenzi, C., et al. 2009. Evaluation of young and adult dogs' reactivity. *Journal of Veterinary Behavior*, 4, 3-10.
- Shan, S., Xu, F. & Brenig, B. 2021. Genome-wide association studies reveal neurological genes for dog herding, predation, temperament, and trainability traits. *Frontiers in Veterinary Science*, 8, 693290.
- Shang, Z., Lv, H., Zhang, M., Duan, L., Wang, S., et al. 2015. Genome-wide haplotype association study identify tnfrsf1a, casp7, lrp1b, cdh1 and tg genes associated with alzheimer's disease in caribbean hispanic individuals. *Oncotarget*, 6, 42504.
- Shao, Z., Noh, H., Bin Kim, W., Ni, P., Nguyen, C., et al. 2019. Dysregulated protocadherin-pathway activity as an intrinsic defect in induced pluripotent stem cell-derived cortical interneurons from subjects with schizophrenia. *Nat Neurosci*, 22, 229-242.
- Shastri, B. S. 2007. Snps in disease gene mapping, medicinal drug development and evolution. *Journal of Human Genetics*, 52, 871-880.
- Sheet, S., Kim, J.-S., Ko, M.-J., Kim, N.-Y., Lim, Y.-J., et al. 2021. Insight into the candidate genes and enriched pathways associated with height, length, length to height ratio and body-weight of korean indigenous breed, jindo dog using gene set enrichment-based gwas analysis. *Animals* [Online], 11.
- Sheppard, G. & Mills, D. 2003. Evaluation of dog-appeasing pheromone as a potential treatment for dogs fearful of fireworks. *Veterinary Record*, 152, 432-436.
- Sherman, B. L. 2008. Separation anxiety in dogs. *Compendium*, 30, 27-42.
- Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., et al. 2022. David: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res*, 50, W216-w221.

- Sherva, R., Zhang, R., Sahelijo, N., Jun, G., Anglin, T., et al. 2023. African ancestry gwas of dementia in a large military cohort identifies significant risk loci. *Mol Psychiatry*, 28, 1293-1302.
- Shettleworth, S. J. 2001. Animal cognition and animal behaviour. *Animal Behaviour*, 61, 277-286.
- Shin, Y.-J. & Shin, N.-S. 2017. Relationship between sociability toward humans and physiological stress in dogs. *Journal of Veterinary Medical Science*, 79, 1278-1283.
- Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., et al. 2012. What is temperament now? Assessing progress in temperament research on the twenty-fifth anniversary of goldsmith et al.(). *Child Development Perspectives*, 6, 436-444.
- Shiner, R. L. & DeYoung, C. G. 2013. The structure of temperament and personality traits: A developmental perspective. *The oxford handbook of developmental psychology, vol. 2: Self and other*. Oxford University Press.
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., et al. 2012. Public participation in scientific research: A framework for deliberate design. *Ecology and society*, 17.
- Sinn, D. L., Gosling, S. D. & Hilliard, S. 2010. Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Applied Animal Behaviour Science*, 127, 51-65.
- Siviter, H., Charles Deeming, D., Rosenberger, J., Burman, O. H., Moszuti, S. A., et al. 2017. The impact of egg incubation temperature on the personality of oviparous reptiles. *Animal Cognition*, 20, 109-116.
- Smith, S. P., Phillips, J. B., Johnson, M. L., Abbot, P., Capra, J. A., et al. 2019. Genome-wide association analysis uncovers variants for reproductive variation across dog breeds and links to domestication. *Evolution, Medicine and Public Health*, 2019, 93-103.
- Soo, M. & Worth, A. 2015. Canine hip dysplasia: Phenotypic scoring and the role of estimated breeding value analysis. *New Zealand Veterinary Journal*, 63, 69-78.
- Speed, D., Hemani, G., Johnson, Michael R. & Balding, David J. 2012. Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics*, 91, 1011-1021.
- St Pourcain, B., Skuse, D. H., Mandy, W. P., Wang, K., Hakonarson, H., et al. 2014. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism*, 5, 1-12.
- Staniland, A. A. & McMahon, S. B. 2009. Mice lacking acid-sensing ion channels (asic) 1 or 2, but not asic3, show increased pain behaviour in the formalin test. *European Journal of Pain*, 13, 554-63.
- Stead, J. D., Clinton, S., Neal, C., Schneider, J., Jama, A., et al. 2006. Selective breeding for divergence in novelty-seeking traits: Heritability and enrichment in spontaneous anxiety-related behaviors. *Behavior Genetics*, 36, 697-712.
- Stellato, A. C., Flint, H. E., Widowski, T. M., Serpell, J. A. & Niel, L. 2017. Assessment of fear-related behaviours displayed by companion

- dogs (*canis familiaris*) in response to social and non-social stimuli. *Applied Animal Behaviour Science*, 188, 84-90.
- Stempiń, W. & Strychalski, J. 2024. A method for assessing dogs in a test evaluating dogs' suitability for animal-assisted education. *Animals*, 14, 1149.
- Stevenson, J., Sonuga-Barke, E., McCann, D., Grimshaw, K., Parker, K. M., et al. 2010. The role of histamine degradation gene polymorphisms in moderating the effects of food additives on children's adhd symptoms. *American Journal of Psychiatry*, 167, 1108-15.
- Stolzechner, L., Bonorand, A. & Riemer, S. 2022. Optimising puppy socialisation—short-and long-term effects of a training programme during the early socialisation period. *Animals*, 12, 3067.
- Sturman, O., von Ziegler, L., Schläppi, C., Akyol, F., Privitera, M., et al. 2020. Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology*, 45, 1942-1952.
- Suarez, L., Bautista-Castaño, I., Peña Romera, C., Montoya-Alonso, J. A. & Corbera, J. A. 2022. Is dog owner obesity a risk factor for canine obesity? A “one-health” study on human–animal interaction in a region with a high prevalence of obesity. *Veterinary Sciences*, 9, 243.
- Sun, Q., Gao, J., An, R., Wang, M. & Wang, Y. 2024. Probing molecular pathways: Illuminating the connection between covid-19 and alzheimer's disease through the endocannabinoid system dynamics. *Journal of Medical Virology*, 96, e29590.
- Sun, R., Wang, Z., Claus Henn, B., Su, L., Lu, Q., et al. 2020. Identification of novel loci associated with infant cognitive ability. *Mol Psychiatry*, 25, 3010-3019.
- Sundman, A. S., Johnsson, M., Wright, D. & Jensen, P. 2016. Similar recent selection criteria associated with different behavioural effects in two dog breeds. *Genes, Brain and Behavior*, 15, 750-756.
- Svartberg, K. 2002. Shyness–boldness predicts performance in working dogs. *Applied Animal Behaviour Science*, 79, 157-174.
- Svartberg, K. & Forkman, B. 2002. Personality traits in the domestic dog (*canis familiaris*). *Applied Animal Behaviour Science*, 79, 133-156.
- Svartberg, K., Tapper, I., Temrin, H., Radesäter, T. & Thorman, S. 2005. Consistency of personality traits in dogs. *Animal Behaviour*, 69, 283-291.
- Svartberg, K. 2006. Breed-typical behaviour in dogs—historical remnants or recent constructs? *Applied Animal Behaviour Science*, 96, 293-313.
- Svartberg, K. 2021. The hierarchical structure of dog personality in a new behavioural assessment: A validation approach. *Applied Animal Behaviour Science*, 238, 105302.
- Svendsen, P. M., Hansen, B. K., Malmkvist, J., Hansen, S. W., Palme, R., et al. 2007. Selection against stereotypic behaviour may have contradictory consequences for the welfare of farm mink (*mustela vison*). *Applied Animal Behaviour Science*, 107, 110-119.

- Svrakic, D. M., Draganic, S., Hill, K., Bayon, C., Przybeck, T., et al. 2002. Temperament, character, and personality disorders: Etiologic, diagnostic, treatment issues. *Acta Psychiatrica Scandinavica*, 106, 189-195.
- Sweeney, P., Bedenbaugh, M. N., Maldonado, J., Pan, P., Fowler, K., et al. 2021. The melanocortin-3 receptor is a pharmacological target for the regulation of anorexia. *Science translational medicine*, 13, eabd6434.
- Swinscow, T. D. V. 1997. Statistics at square one. Available: <https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/preface> [Accessed 04 January 2021].
- Sydney, U. o. 2023. *Online mendelian inheritance in animals* [Online]. Available: <https://omia.org/home/> [Accessed 20/01/2023 2023].
- Tabakhnick, B. & Fidell, L. S. 2007. *Using multivariate statistics*, University of Michigan, Allyn and Bacon, 2001.
- Taber, K. S. 2018. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in science education*, 48, 1273-1296.
- Takeuchi, Y., Hashizume, C., Arata, S., Inoue-Murayama, M., Maki, T., et al. 2009a. An approach to canine behavioural genetics employing guide dogs for the blind. *Animal Genetics*, 40, 217-24.
- Takeuchi, Y., Kaneko, F., Hashizume, C., Masuda, K., Ogata, N., et al. 2009b. Association analysis between canine behavioural traits and genetic polymorphisms in the shiba inu breed. *Animal Genetics*, 40, 616-622.
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., et al. 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20, 467-484.
- Tami, G. & Gallagher, A. 2009. Description of the behaviour of domestic dog (*canis familiaris*) by experienced and inexperienced people. *Applied Animal Behaviour Science*, 120, 159-169.
- Taş, B. 2025. Rethinking love, independence, and speciesism in assistance dog discourse. *Frontiers in Sociology*, 9, 1448676.
- Taşkıran, E. Z., Karaosmanoğlu, B., Koşukcu, C., Ürel-Demir, G., Akgün-Doğan, Ö., et al. 2021. Diagnostic yield of whole-exome sequencing in non-syndromic intellectual disability. *Journal of Intellectual Disability Research*, 65, 577-588.
- Tavakol, M. & Dennick, R. 2011. Making sense of cronbach's alpha. *International journal of medical education*, 2, 53.
- Tavakol, M. & Wetzal, A. 2020. Factor analysis: A means for theory and instrument development in support of construct validity. *International journal of medical education*, 11, 245.
- Taylor, K. D. & Mills, D. S. 2006. The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior*, 1, 94-108.
- Temesi, A., Turcsán, B. & Miklósi, Á. 2014. Measuring fear in dogs by questionnaires: An exploratory study toward a standardized inventory. *Applied Animal Behaviour Science*, 161, 121-130.

- Teng, X., Aouacheria, A., Lionnard, L., Metz, K. A., Soane, L., et al. 2019. Kctd: A new gene family involved in neurodevelopmental and neuropsychiatric disorders. *CNS Neuroscience and Therapeutics*, 25, 887-902.
- Thornton, A. & Truskanov, N. 2022. The role of natural history in animal cognition. *Current Opinion in Behavioral Sciences*, 46, 101154.
- Thorsrud, J. A., Evans, K. M., Quigley, K. C., Srikanth, K. & Huson, H. J. 2025. Performance comparison of genomic best linear unbiased prediction and four machine learning models for estimating genomic breeding values in working dogs. *Animals*, 15, 408.
- Tiira, K. & Lohi, H. 2014. Reliability and validity of a questionnaire survey in canine anxiety research. *Applied Animal Behaviour Science*, 155, 82-92.
- Tiira, K. & Lohi, H. 2015. Early life experiences and exercise associate with canine anxieties. *PloS one*, 10, e0141907.
- Tomida, S., Mamiya, T., Sakamaki, H., Miura, M., Aosaki, T., et al. 2009. Usp46 is a quantitative trait gene regulating mouse immobile behavior in the tail suspension and forced swimming tests. *Nature Genetics*, 41, 688-695.
- Tomkins, L. M., Thomson, P. C. & McGreevy, P. D. 2011. Lateralization and temperament tests can predict guide dog success. *Journal of Veterinary Behavior*, 6, 89.
- Topál, J., Miklósi, Á. & Csányi, V. 1997. Dog–human relationship affects problem-solving behavior in the dog. *Anthrozoös*, 10, 214-224.
- Trubetsky, V., Pardiñas, A. F., Qi, T., Panagiotaropoulou, G., Awasthi, S., et al. 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*, 604, 502-508.
- Turcsán, B., Kubinyi, E. & Miklósi, Á. 2011. Trainability and boldness traits differ between dog breed clusters based on conventional breed categories and genetic relatedness. *Applied Animal Behaviour Science*, 132, 61-70.
- Turcsán, B., Range, F., Virányi, Z., Miklósi, Á. & Kubinyi, E. 2012. Birds of a feather flock together? Perceived personality matching in owner–dog dyads. *Applied Animal Behaviour Science*, 140, 154-160.
- UFAW. 2024. *Genetic welfare problems of companion animals* [Online]. Available: <https://www.ufaw.org.uk/dogs/dogs> [Accessed 25 January 2025].
- Vallat, R. 2018. Pingouin: Statistics in python. *The Journal of Open Source Software*, 3, 1026.
- Valletta, J. J., Torney, C., Kings, M., Thornton, A. & Madden, J. 2017. Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203-220.
- Valsecchi, P., Previde, E. P., Accorsi, P. A. & Fallani, G. 2010. Development of the attachment bond in guide dogs. *Applied Animal Behaviour Science*, 123, 43-50.
- Valsecchi, P., Barnard, S., Stefanini, C. & Normando, S. 2011. Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment. *Journal of Veterinary Behavior*, 6, 161-177.

- van den Berg, L., Schilder, M. B., de Vries, H., Leegwater, P. A. & van Oost, B. A. 2006. Phenotyping of aggressive behavior in golden retriever dogs with a questionnaire. *Behavioural genetics*, 36, 882-902.
- van der Borg, J. A., Graat, E. A. & Beerda, B. 2017. Behavioural testing based breeding policy reduces the prevalence of fear and aggression related behaviour in rottweilers. *Applied Animal Behaviour Science*, 195, 80-86.
- Van der Waaij, E., Wilsson, E. & Strandberg, E. 2008. Genetic analysis of results of a swedish behavior test on german shepherd dogs and labrador retrievers. *Journal of Animal Science*, 86, 2853-2861.
- Van Griethuijsen, R. A., van Eijck, M. W., Haste, H., Den Brok, P. J., Skinner, N. C., et al. 2015. Global patterns in students' views of science and interest in science. *Research in science education*, 45, 581-603.
- Van Steenbeek, F., Hytönen, M., Leegwater, P. & Lohi, H. 2016. The canine era: The rise of a biomedical model. *Animal Genetics*, 47, 519-527.
- Vaterlaws-Whiteside, H. & Hartmann, A. 2017. Improving puppy behavior using a new standardized socialization program. *Applied Animal Behaviour Science*, 197, 55-61.
- Veerkamp, R. F., Mulder, H. A., Thompson, R. & Calus, M. P. L. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *Journal of Dairy Science*, 94, 4189-4197.
- Verhoeven, W. M. A., Egger, J. I. M., Janssen, P. K. C. & van Haeringen, A. 2020. Adult male patient with severe intellectual disability caused by a homozygous mutation in the hnmt gene. *BMJ Case Rep*, 13.
- Verma, R., Holmans, P., Knowles, J. A., Grover, D., Evgrafov, O. V., et al. 2008. Linkage disequilibrium mapping of a chromosome 15q25-26 major depression linkage region and sequencing of ntrk3. *Biol Psychiatry*, 63, 1185-9.
- Vignal, A., Milan, D., SanCristobal, M. & Eggen, A. 2002. A review on snp and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34, 275-305.
- Visscher, P. M., Hill, W. G. & Wray, N. R. 2008. Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics*, 9, 255-266.
- Visscher, P. M., Hemani, G., Vinkhuyzen, A. A., Chen, G.-B., Lee, S. H., et al. 2014. Statistical power to detect genetic (co) variance of complex traits using snp data in unrelated samples. *PLoS Genetics*, 10, e1004269.
- Völter, C. J., Starić, D. & Huber, L. 2023. Using machine learning to track dogs' exploratory behaviour in the presence and absence of their caregiver. *Animal Behaviour*, 197, 97-111.
- von Ziegler, L., Sturman, O. & Bohacek, J. 2021. Big behavior: Challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology*, 46, 33-44.
- Wallis, L. J., Szabó, D. & Kubinyi, E. 2020. Cross-sectional age differences in canine personality traits; influence of breed, sex,

- previous trauma, and dog obedience tasks. *Frontiers in Veterinary Science*, 6, 493.
- Wallis, N. J., Sumanasekera, N. T. & Raffan, E. 2024. Obesity risk factors in british labrador retrievers: Effect of sex, neuter status, age, chocolate coat colour and food motivation. *Veterinary Record*, 194, no-no.
- Wang, H., Atif, O., Tian, J., Lee, J., Park, D., et al. 2022. Multi-level hierarchical complex behavior monitoring system for dog psychological separation anxiety symptoms. *Sensors (Basel)*, 22.
- Wang, K.-S., Liu, X.-F. & Aragam, N. 2010. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophrenia research*, 124, 192-199.
- Wang, K., Zhang, H., Bloss, C., Duvvuri, V., Kaye, W., et al. 2011. A genome-wide association study on common snps and rare cnvs in anorexia nervosa. *Mol Psychiatry*, 16, 949-959.
- Wang, S., Strandberg, E., Arvelius, P., Clements, D. N., Wiener, P., et al. 2021. Genome-wide association studies for canine hip dysplasia in single and multiple populations – implications and potential novel risk loci. *BMC Genomics*, 22, 636.
- Wang, T., Hoekzema, K., Vecchio, D., Wu, H., Sulovari, A., et al. 2020. Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nature communications*, 11, 4932.
- Wanner, C. 2017. 'For the good of the breed': Care, ethics, and responsibility in pedigree dog breeding.
- Wayne, R. K. & Ostrander, E. A. 1999. Origin, genetic diversity, and genome structure of the domestic dog. *BioEssays*, 21, 247-257.
- Werren, E. A., Rodriguez Bey, G., Majethia, P., Kaur, P., Patil, S. J., et al. 2024. Biallelic epb41l3 variants underlie a developmental disorder with seizures and myelination defects. *Brain*, awae299.
- Westgarth, C., Brooke, M. & Christley, R. M. 2018. How many people have been bitten by dogs? A cross-sectional survey of prevalence, incidence and factors associated with dog bites in a uk community. *Journal of Epidemiology and Community Health*, 72, 331-336.
- Westgarth, C., Christley, R. M., Marvin, G. & Perkins, E. 2019. The responsible dog owner: The construction of responsibility. *Anthrozoös*, 32, 631-646.
- Whiteside, H. 2017. Service dog temperament versus behaviour: What to measure and when. Available: <https://summer2017.iaabcjournal.org/service-dog-temperament/> [Accessed 06 August 2020].
- Whitmarsh, L. 2005. The benefits of guide dog ownership. *Visual Impairment Research*, 7, 27-42.
- Wiener, P., Ilska, J., Lofgren, S., Sanchez-Molano, E., Clements, D., et al. 2015. *Dissecting genetic and non-genetic influences on dog personality*.
- Wiener, P. & Haskell, M. J. 2016. Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior*, 16, 81-85.
- Wiklund-Hörnqvist, C., Jonsson, B. & Nyberg, L. 2014. Strengthening concept learning by repeated testing. *Scandinavian journal of psychology*, 55, 10-16.

- Williams, E., Cleghern, Z., Foster, M., Holder, T., Roberts, D., et al. A smart collar for assessment of activity levels and environmental conditions for guide dogs. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020. IEEE, 4628-4631.
- Williams, M. J., Goergen, P., Phad, G., Fredriksson, R. & Schiöth, H. B. 2014. The drosophila kctd-family homologue kctd12-like modulates male aggression and mating behaviour. *European Journal of Neuroscience*, 40, 2513-26.
- Willis, M. B. 1989. *Genetics of the dog*, New York, Howell Book House.
- Wilsson, E. & Sundgren, P.-E. 1997. The use of a behaviour test for the selection of dogs for service and breeding, i: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Applied Animal Behaviour Science*, 53, 279-295.
- Wilsson, E. & Sundgren, P.-E. 1998. Behaviour test for eight-week old puppies—heritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*, 58, 151-162.
- Wirth, K. E. & Rein, D. B. 2008. The economic costs and benefits of dog guides for the blind. *Ophthalmic Epidemiology*, 15, 92-98.
- Wood, C. D. 2020. Assessing the role personality plays in puppy raisers and guide dogs in training.
- Working Dog Project. n.d. *The working dog project* [Online]. Available: <https://workingdogproject.org/> [Accessed 07 November 2020].
- Wormald, D., Lawrence, A. J., Carter, G. & Fisher, A. D. 2016. Validation of modified open field behaviour as a measure of trait anxiety in the dog. *Applied Animal Behaviour Science*, 179, 95-102.
- Wright, C., Shin, J. H., Rajpurohit, A., Deep-Soboslay, A., Collado-Torres, L., et al. 2017. Altered expression of histamine signaling genes in autism spectrum disorder. *Transl Psychiatry*, 7, e1126.
- Wright, H. F., Mills, D. S. & Pollux, P. M. 2012. Behavioural and physiological correlates of impulsivity in the domestic dog (canis familiaris). *Physiology and behavior*, 105, 676-682.
- Wu, Y., Cao, H., Baranova, A., Huang, H., Li, S., et al. 2020. Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Translational Psychiatry*, 10, 209.
- Xia, C., Amador, C., Huffman, J., Trochet, H., Campbell, A., et al. 2016. Pedigree- and snp-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. *PLoS Genetics*, 12, e1005804.
- Xue, A., Jiang, L., Zhu, Z., Wray, N. R., Visscher, P. M., et al. 2021. Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. *Nature communications*, 12, 1-11.
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. 2011. Gcta: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88, 76-82.

- Yilmaz, Z., Halvorsen, M., Bryois, J., Yu, D., Thornton, L. M., et al. 2020. Examination of the shared genetic basis of anorexia nervosa and obsessive–compulsive disorder. *Mol Psychiatry*, 25, 2036-2046.
- York, S. & Whiteside, H. 2018. Exploring social (non-working) behavior in guide dogs from the perspective of guide dog owners. *Anthrozoös*, 31, 525-536.
- Yu, D., Tan, Y., Chakraborty, M., Tomchik, S. & Davis, R. L. 2018. Elongator complex is required for long-term olfactory memory formation in drosophila. *Learning and Memory*, 25, 183-196.
- Zapata, I., Serpell, J. A. & Alvarez, C. E. 2016. Genetic mapping of canine fear and aggression. *BMC Genomics*, 17, 572.
- Zeigler-Hill, V. & Highfill, L. 2010. Applying the interpersonal circumplex to the behavioral styles of dogs and cats. *Applied Animal Behaviour Science*, 124, 104-112.
- Zheng, Y., Verhoeff, T. A., Perez Pardo, P., Garssen, J. & Kraneveld, A. D. 2020. The gut-brain axis in autism spectrum disorder: A focus on the metalloproteases adam10 and adam17. *Int J Mol Sci*, 22.
- Zhu, H. & Zhou, X. 2020. Statistical methods for snp heritability estimation and partition: A review. *Computational and Structural Biotechnology Journal*, 18, 1557-1568.