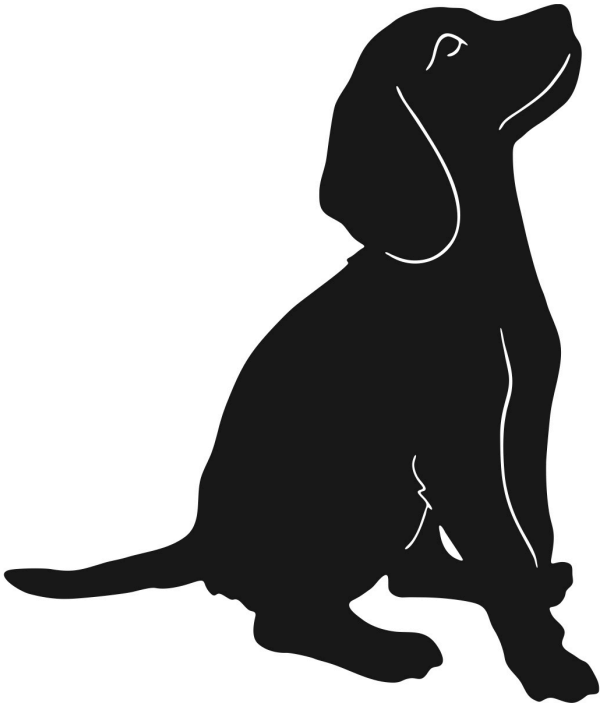
University of Stirling

A practical framework for harmonising welfare and quality of data output in the laboratory-housed dog

Laura E. Hall

Psychology
School of Natural Sciences

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy



August 2014

Presentations given

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2014) *Refining oral gavage: Assessing and improving welfare in the laboratory-housed dog.* Poster presentation at 9th World Congress on Alternatives to Animals in the Life Sciences, Prague, August 2014.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2014) *Refining oral gavage: Assessing and improving welfare in the laboratory-housed dog.* Oral presentation at UFAW meeting, York, June 2014.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2014) *Refining oral gavage: Assessing and improving welfare in the laboratory-housed dog.* Oral presentation at industry facility, April 2014.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2014) *A practical guide to welfare assessment and improvement in the laboratory-housed dog.* Oral presentation at IAT Congress, UK, April 2014.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *Refining oral gavage: Assessing and improving welfare in the laboratory-housed dog.* Invited oral presentation at LASA Winter Meeting, London, November 2013.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Poster presented at LASA Winter Meeting, London, November 2013.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Internal CPD presentation, Alderley Park, November 2013.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Internal CPD presentation, Alderley Park, November 2013.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *An introduction to positive reinforcement training for the laboratory-housed dog.* Internal CPD presentation, Alderley Park, November 2013.
- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Oral presentation at Animal Concepts, Edinburgh, September 2013.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Poster presented at FELASA, Barcelona, June 2013.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2013) *A Welfare Assessment Framework for laboratory-housed dogs: linking welfare and quality of data output.* Oral presentation, Behaviour and Evolution Research Group meeting, Psychology, University of Stirling, May 2013.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2012) *Development of a Welfare Assessment Framework for laboratory-housed beagles: affect, behaviour and cardiovascular analysis.* Invited oral presentation at LASA Winter Meeting, Manchester, November 2012.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2012) *Development of a Welfare Assessment Framework for laboratory-housed beagles.* Internal oral presentation, Alderley Park, July 2012.

- Hall, L.E., Robinson, S., Moors, J. and Buchanan-Smith, H.M. (2012) *Development of a Welfare Assessment Framework for laboratory-housed beagles: affect, behaviour and cardiovascular analysis.* Poster presented at ASAB/SEB/NC3Rs meeting, London, June 2012.

- Hall, L.E., Robinson, S., Moors, J. and Buchanan-Smith, H.M. (2012) *Developing a Welfare Assessment Framework for laboratory-housed beagles: affect, behaviour and cardiovascular analysis.* Poster presentation at UFAW meeting, York, June 2012.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2011) *Refinements in dog housing and husbandry, and the link with quality of scientific output.* Poster presented at Christmas Poster Competition, Psychology, University of Stirling, December 2012.

- Hall, L.E., Robinson, S. and Buchanan-Smith, H.M. (2011) *Refinements in dog housing and husbandry, and the link with quality of scientific output.* Poster presented at 8th World Congress on Alternatives to Animals in the Life Sciences, Montreal, August 2011.

"It is necessary to help others, not only in our prayers, but in our daily lives. If we find we cannot help others, the least we can do is to desist from harming them"

HH XXIV Dalai Lama

Contents

Abstract

In the UK, laboratory-housed dogs are primarily used as a non-rodent species in the safety testing of new medicines and other chemical entities. The use of animals in research is governed by the Animals (Scientific Procedures) Act (1986, amended 2012) and legislation is underpinned by the principles of humane experimental technique: Replacement, Reduction and Refinement. A link between animal welfare and the quality of data produced has been shown in other species (e.g. rodents, nonhuman primates), however, no established, integrated methodology for identifying or monitoring welfare and quality of data output previously existed in the laboratory-housed dog. In order to investigate the effects of planned Refinements to various aspects of husbandry and regulated procedures, this project sought to integrate behavioural, physiological and other measures (e.g. cognitive bias, mechanical pressure threshold) and to provide a means for staff to monitor welfare whilst also establishing the relationship between welfare and quality of data output.

Affective state was identified using an established method of cognitive bias testing, before measuring welfare at 'baseline' using measures of behaviour and physiology. Dogs then underwent 'positive' and 'negative' behavioural challenges to identify the measures most sensitive to changing welfare and most suitable for use in a framework. The resulting Welfare Assessment Framework, developed in three groups of dogs from contrasting backgrounds within the facility, found a consistent pattern of behaviour, cardiovascular function, affect and mechanical pressure threshold (MPT). Dogs with a negative affective state had higher blood pressure at baseline than those with positive affective states, and the magnitude of the effect of negative welfare suggests that welfare may act as a confound in the interpretation of cardiovascular data.

The responses to restraint included increases in blood pressure and heart rate measures which approached ceiling levels, potentially reducing the sensitivity of measurement. If maintained over time this response could potentially have a negative health impact on other organ systems and affecting the data obtained from those. Dogs with a negative welfare state also had a lower mechanical pressure threshold, meaning they potentially experienced greater stimulation from unpleasant physical stimuli. Taken together with the behaviours associated with a negative welfare state (predominantly vigilant or stereotypic behaviours) the data suggest that dogs with a negative welfare state have a greater behavioural and physiological response to stimuli in their environment; as such, data obtained from their use is different from that obtained from dogs with a positive welfare state. This was confirmed by examining the effect size (Cohen's d) resulting from the analysis of affective state on cardiovascular data. An increase in variance, particularly in the small dog numbers typical of safety assessment studies, means a reduction in the power of the study to detect the effect under observation; a decrease in variation has the potential to reduce the number of dogs use, in line with the principle of Reduction and good scientific practice.

The development of the framework also identified areas of the laboratory environment suitable for Refinement (e.g. restriction to single-housing and restraint) and other easily-implemented Refinements (e.g. feeding toy and human interaction) which could be used to improve welfare. As a result of this, a Welfare Monitoring Tool (WMT) in the form of a tick sheet was developed for technical and scientific staff to identify those dogs at risk of reduced welfare and producing poor quality data, as well as to monitor the effects of Refinements to protocols.

Oral gavage is a common regulated procedure, known to be potentially aversive and was identified as an area in need of Refinement. A program of desensitisation and

positive reinforcement training was implemented in a study also comparing the effects of a sham dose condition versus a control, no-training, condition. A number of the measures used, including home pen behaviour, behaviour during dosing, MPT and the WMT showed significant benefits to the dogs in the Refined condition. Conversely, dogs in the sham dose condition showed more signs of distress and took longer to dose than dogs in the control condition. The welfare of control dogs was intermediate to sham dose and Refined protocol dogs.

This project identified a positive relationship between positive welfare and higher quality of data output. It developed and validated a practical and feasible means of measuring welfare in the laboratory environment in the Welfare Assessment Framework, identified areas in need of Refinement and developed practical ways to implement such Refinements to husbandry and regulated procedures. As such it should have wide implications for the pharmaceutical industry and other users of dogs in scientific research.

Acknowledgements

There are so many people who have contributed to the success of this project, which reflects the enthusiasm with which it has been embraced and the conviction of those involved. Both this project and I have thrived in an environment in which so many people have strived for improvement.

My primary supervisor, Hannah Buchanan-Smith, has been tremendously patient and supportive, both in sharing her immense knowledge and in always ensuring that the project would be as successful as possible. Hannah has challenged me and supported me throughout the project and I have immensely enjoyed being able to indulge my passion for this research.

Similarly, my industrial supervisor Dr Sally Robinson gave invaluable support while I was on-site, ensured that I was able to experience all aspects of research and see my project in context, and provided me with opportunities to travel to and present at conferences. Moreover, Sally instigated and has supported this project beyond the scope of the thesis. My second supervisor, Dr Christine Caldwell has provided encouragement and feedback throughout the project. Stewart D and Mary helped in understanding how things worked in an unfamiliar environment; Jackie, Matt and Nicki provided assistance with capturing and analysing cardiovascular data; Stewart B provided excellent assistance in data collection for the final study, and thanks also to Lindsey for being so passionate in ensuring that the findings of the project were implemented. CD, EW and KH exceeded my expectations in their ability to implement my training protocols and incorporated many extra duties into their already hectic schedules. GSK have also shared their knowledge and enthusiasm with me.

Dr Lou Tasker provided invaluable input to the project and her expertise and moral support contributed greatly to the completion of this thesis. Dr Debra Lynn and Dr Kris Descovich provided me with proof-reading help, shared their knowledge and much moral support in the final stages of my write-up. I am also grateful to the support staff and technicians in Psychology who have also provided support throughout the project. Statistical assistance was provided by Dr Peter Cahusac, Dr Kate Howie and Dr Ian Peers. Dr Mark Prescott, Dr Robert Hubrecht and Dr Daniel Mills kindly provided insight in the early stages of the project.

This project was generously funded by a BBSRC Industrial CASE Studentship and AstraZeneca. I received funding to travel to and present at conferences from the 8th and 9th World Congresses on Alternatives, FELASA, LASA, NC3Rs and the Psychology division.

I am most grateful to my family and friends, without whom I would not have been able to complete this project. My mother engendered in me an interest in animal welfare from an early age and has undoubtedly provided me with my curiosity and stubborn determination. Thomas has been there through the highs, lows and tears, and has read this thesis through more times than anyone should have to! My family, and adopted family, have made sure I managed to get here. And my friends have been patient through the long absences during data collection and deeply immersed in the write up. Anna frequently provided very welcome distraction and moral support. Jade and Fly deserve much credit for putting up with upheaval in their lives without complaint; Jade has always been the model of calmness and resilience, and not many dogs can have spent their lives travelling back and forth on a train as much as Fly. She has well earned her status as PhD Support Puppy!

My final acknowledgements go to the 40 dogs who participated in this project, for their unfailing ability to be happy to see me in the morning, for their enthusiastic acceptance of my presence and for providing me with invaluable learning experiences. This is for them, and for all other laboratory-housed dogs, they have contributed more than they know.

Glossary of terms used in the thesis

Definitions

Welfare *An individual's state in relation to its attempts to cope with a situation, both in terms of physical health and subjective experience* Donald Broom, 1986

Affective state *Free-floating mood states, not directed at an object, requiring a lesser degree of information processing* Elizabeth Paul and colleagues, 2005

Stress *Stress is a broad term for specific morphological, biochemical, physiological or behavioural changes experienced by an organism in response to a stressful event or stressor* Wolfgang H. Vogel, 1987

Distress *The result of exposure to stressors in the environment which over-taxes an individual's coping systems and reduces, or seems likely to reduce, its fitness* Donald Broom, 1986

Suffering *The internal emotional state which results from chronic overloading of the coping mechanisms* Donald Broom, 1998

Environmental enrichment *An environmental or social provision which seeks to improve welfare and/or encourage the display of a natural repertoire of behaviours, which can be conditionally beneficial or beneficial to all subjects* Robert Hubrecht, 2010

Replacement *Refers to methods which avoid or replace the use of animals defined as "protected" under the UK Animal (Scientific Procedures) Act 1986. These can be absolute replacements (e.g. computer modelling, in vitro methods, human volunteers) or relative replacements (e.g. invertebrates, such as fruit flies and nematode worms)* NC3Rs operational definition

Reduction *Refers to methods which minimise animal use and enable researchers to obtain comparable levels of information from fewer animals or to obtain more information from the same number of animals, thereby reducing future use of animals (e.g. improved experimental design, modern imaging techniques, sharing data and resources)* NC3Rs operational definition

Refinement *Any approach which avoids or minimises the actual or potential pain, distress and other adverse effects experienced at any time during the life of the animals involved, and which enhances their wellbeing* Hannah Buchanan-Smith and colleagues, 2005

Abbreviations

NC3Rs The National Centre for the 3Rs; the body responsible for promoting the uptake of the 3Rs in the UK.

PAS *Positive affective state*

NAS *Negative affective state*

SP *Safety Pharmacology group of dogs* (*long-term, intensive-use colony*)

DMPK '*Dose molecular pharmacokinetic' study group of dogs (long-term, low use colony*

Stock *Naïve group of dogs, held as stock for upcoming studies*

SD *Sham dosing, a protocol used to habituate dogs to dosing in the pre-study phase*

RP *Refined protocol, a protocol devised and implemented for dosing in Chapter 7, with the aim of reducing the negative welfare impact*

MPT Mechanical Pressure Threshold

MPTT Mechanical Pressure Threshold Testing

NWI Negative Welfare Indicator

PWI Positive Welfare Indicator

↑ *An increase in a parameter which indicates a positive change in welfare*

↑ *An increase in a parameter which indicates a negative change in welfare*

↓ *A decrease in a parameter which indicates a positive change in welfare*

↓ *A decrease in a parameter which indicates a negative change in welfare*

X or ↑ *or* ↓ *A change in a parameter which is context-specific or welfare-neutral*

CHAPTER 1

The use of the dog in laboratories for research and testing

"A bustling, eager little dog, full of enthusiasm and vigour, ever ready for any activity that involves him. Sturdy, bold and active, he is the very essence of quality, and is blessed with an equable and merry temperament"

The Beagle, Kennel Club Breed Standards (2011)

Abstract

This chapter lays out the background to and premise of the project. In the UK, dogs are primarily used in the safety assessment of new medicines and their use is governed by the Animals (Scientific Procedures) Act (1986, amended 2012), a transposition of European Directive 2010/63/EU on the protection of animals used for scientific purposes. Regulated procedures, which have the potential to cause pain, suffering, distress or lasting harm, also have the potential to negatively impact on welfare. The 3Rs (Replacement, Reduction, Refinement) are the principles on which humane science is based and are used to guide ethical and scientific decision-making and policy to reduce suffering and promote better quality of scientific practice. The science of animal welfare assessment is discussed, and methods of assessment described. An understanding of the dog's natural

history is also presented to understand its needs. The factors necessary to integrate a comprehensive welfare assessment are presented and their use in this thesis outlined.

1.1 Introduction

Animals have been used as experimental models in human medicines since prehistory (Gad, 2006). Globally, dogs are used for a variety of purposes, whilst in the UK their principal use is to fulfil the legal requirements for the safety testing of new medicines prior to human exposure. There are two crucial reasons to ensure the most humane use of dogs in laboratory settings: our ethical obligation to prevent suffering in a species which experiences pain, discomfort or distress; and our scientific need to ensure that they are fit for use, by which we mean they are valid, reliable and predictive models for safety and efficacy testing of chemicals prior to human use. Legislative (e.g. European Directive 2010/63) and ethical (e.g. the 3Rs) guidelines provide frameworks within which dogs can be used in laboratories. However there remains a paucity of quantitative data on best practice. Research into the natural history of the dog and its welfare are critical to the development and implementation of effective Refinements, methods of minimising the negative impact of the laboratory environment, and promoting positive welfare.

1.1.1 Legislation

The use of live animals in scientific research has long been regulated in the UK. In 1876, the Cruelty to Animals Act became the first law to regulate scientific procedures, which was in force until 1986. At the outset of this project, the governing legislation was presented in the Animals (Scientific Procedures) Act 1986 (A(SP)A), a transposition of the European Directive (86/609/EEC). The Act is supported by associated Codes of Practice (e.g. Home Office, 1989, 1995). The scope of the Act is to regulate the use of animals for scientific purposes; it does so by regulating the procedures which can be performed on protected animals. The definition of a regulated procedure as one which has the *potential* to "cause pain, suffering, distress or lasting harm" (Home Office, 1986), along with "degrees" of severity (mild, moderate or severe) in procedures has become a key feature in the regulation of animal use. In addition, the persons responsible for the design and conduct out of study protocols were defined and minimum training requirements were introduced by A(SP)A to ensuring that
defined standards are adhered to. Personal, project and establishment licenses are administered by the Home Office and must be obtained before regulated procedures and the terms complied with. Under A(SP)A, (1986) all vertebrates beyond half-way

through gestation, and the common octopus, *Octopus vulgaris*, are considered "protected". Four groups of animals (primate, dog, cat and horse) are afforded special protection. Special dispensation and additional justification are required to use an animal given special protection. The 1986 Act incorporated features of humane experimental technique, such as Russell and Burch's 3Rs - Replacement, Reduction and Refinement (Russell & Burch, 1959).

In 2010, a new European Directive (2010/63/EU) was passed (European Union, 2010), updating existing legislation, to harmonise animal use within the EU. It incorporated and promoted more fully the principles of humane experimental technique. The Directive was transposed into UK legislation as an update to A(SP)A in 2012. There are a number of distinct differences between the previous legislation and the new Directive, for example the promotion of positive animal welfare and explicit by reference to the 3Rs. All aspects of animal science including education and fundamental research are included and files must be kept on nonhuman primates, dogs and cats that contain information on personal histories, covering all aspects of the individual (e.g. previoussocial information). Moreover an assessment of the actual severity (Hartung, 2010) rather than anticipated severity is performed under the revised legislation and recorded for publication in statistical information.

Under Directive 2010/63/EU, the need to obtain dogs from designated breeders now also extends to other countries within the EU . This had been the case in the UK under A(SP)A 1986, however, in other, non-EU countries it remains acceptable to use other sources of dogs. For example, in the US as well as suppliers of purpose bred dogs, there are suppliers of random source dogs pedigree (breed or mongrel) which may be obtained from various sources, including shelters. Concerns for animal welfare and public perception of scientific research led to the regulation of the sources of animals for scientific purposes in the UK. Because of the status of dogs as a companion animal, as well as the perceived welfare costs of using non-purpose bred dogs in research, dogs must be obtained from designated breeders, which breed animals specifically for science. Designated breeders are covered by their own Codes of Practice (Home Office, 1995), with standards being broadly similar to those demanded of the laboratory environment. When considering the need to prepare dogs for a working life in a scientific study (Meunier, 2006), it is unacceptable to use dogs sourced from non-purpose bred colonies which have received insufficient or no training.

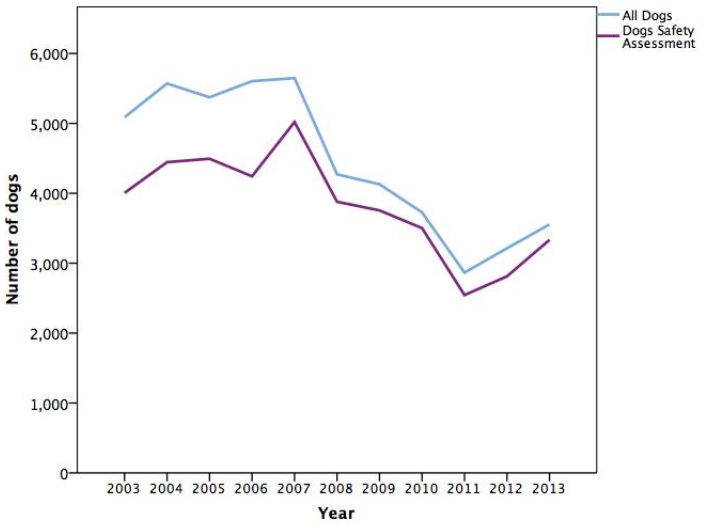


FIGURE 1.1: Total number of dogs used and dogs used in safety assessment per year in Great Britain, from 2003-2013

1.1.2 The use of dogs in laboratories in Great Britain

Annual figures on the number and nature of scientific procedures for Great Britain are published by the Home Office. Overall, the proportion of dogs used compared to other species in comparatively small (e.g. 3,554 dogs used in 2013 of approximately 4 million animals). In the UK, the primary use of dogs is in safety assessment of new therapeutic medicines (Figures 1.1 and 1.2). Other uses, such as fundamental and veterinary research, constitute a smaller proportion of use. Note that a dog can be used in more than one procedure, subject to conditions on re-use, under which dogs subject to mild or short-lasting procedures may be used in additional studies. There is an overall decreasing trend in dog use, with a notable dip in dog numbers from 2007 following the global economic crisis.

1.1.3 The use of dogs in laboratories globally

Dog use has shown an overall decreasing trend in Great Britain, however this is not reflective of global dog use. It is not possible to accurately estimate global dog use due to the variation in data reporting between countries. Unlike Great Britain, many countries do not report the number of animals used annually, while others may do so less frequently or with less detail; an estimated 79% of countries provide *no* data on animal use (Taylor, Gordon, Langley & Higgins, 2008). Not only does this make it difficult to monitor trends in animals use and welfare, but it is also difficult to track the implementation of the 3Rs, in particular a Reduction in animal numbers. Figure

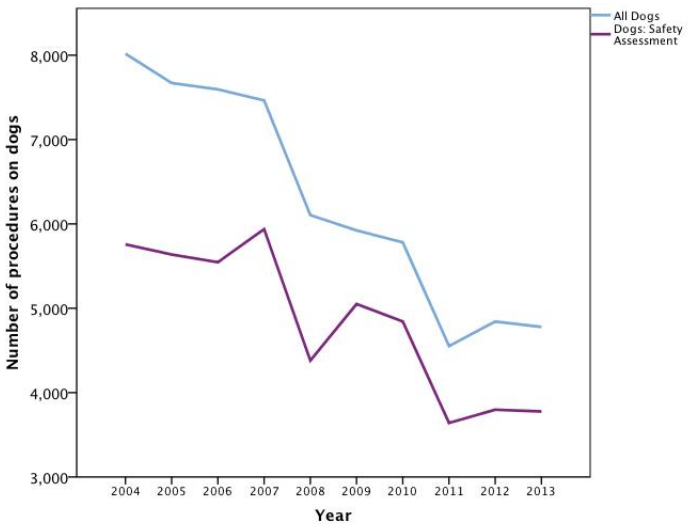


FIGURE 1.2: Total number of procedures and safety assessment procedures on dogs per year in Great Britain, from 2003-2013

1.3 provides an estimate of the extent of global dog use. It is worth noting that there are no data available for countries such as China where animal use is known to be among the highest globally (Taylor et al., 2008). Therefore the data presented in 1.3 should be taken as a very conservative estimate. What is apparent is that the USA is the leading user of dogs (amongst those who report numbers) using over 64,000 dogs, followed by the EU which uses over 21,000 dogs (including Great Britain).

While in the past many end users in the UK may have bred their own dogs for scientific procedures, this practice is diminishing and dogs are predominantly acquired from a small number of designated commercial breeders. Therefore the majority of dogs will be transported to the designated establishment and must undergo a period of acclimatisation, before being used in regulated procedures (Boxall, Heath, Bate & Brautigam, 2004). The lifetime experiences of dogs contribute to their wellbeing and resilience, and Refinement must be applied to all stages of life.

1.1.4 Numbers of dogs used in scientific procedures

It is difficult to estimate the number of dogs used in breeding and scientific procedures for any given pharmaceutical product, particularly as many companies are multi-national and the studies used to develop any given product may be conducted in multiple countries. Phillips et al. (2004) estimated that between 150-290 dogs are used per drug development project, based on data from 10 projects and four companies. However, the authors state that this is likely to be an underestimate as many preliminary studies will not result in a marketed product, and figures from preliminary

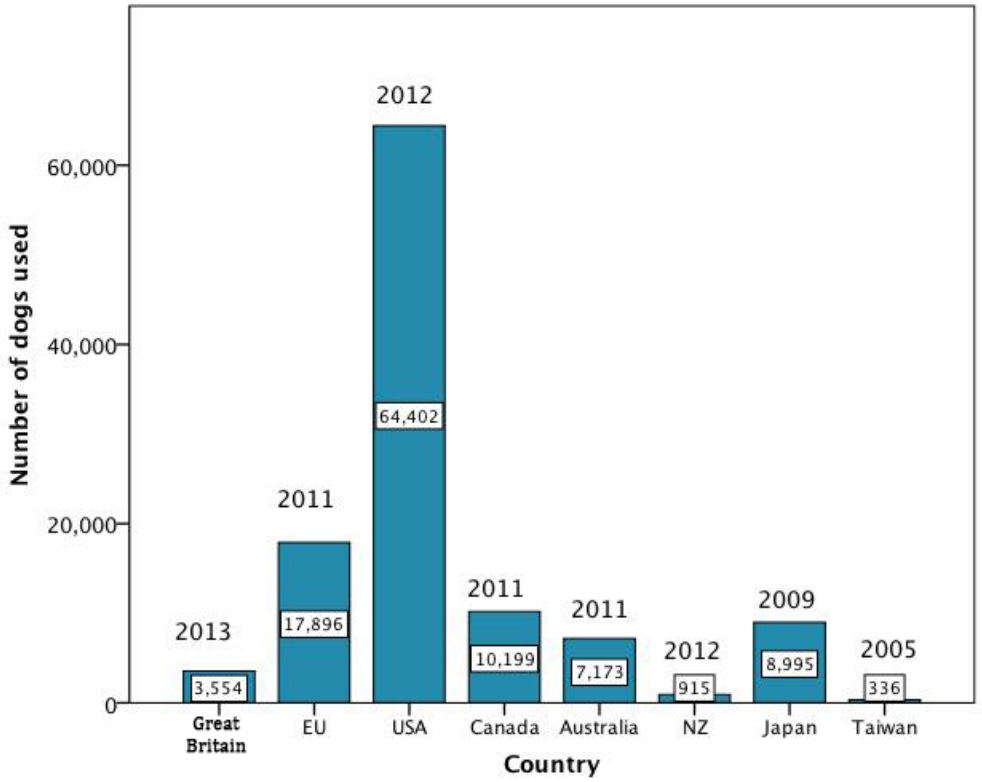


FIGURE 1.3: Global estimates of dog use using most recent figures available; Great Britain: Home Office (2014); EU: European Union (2012); United States of America: USDA (2013); Canada: CCAC (2012); Australia: Animal Research Review Panel NSW (2012); Department of Primary Industries Victoria (2012); Department of Primary Industries, Parks, Water and Environment Tasmania (2012); New Zealand: NAEAC (2013); Japan: Yagami et al (2010); Taiwan: Chen (2007).

studies are not always reported, so the true figure is likely to be much higher. Table 1.1 depicts typical dog use across types of test. Non-clinical testing typically progresses from small discovery studies, to maximum tolerated dose (MTD) or dose range finding (DFF) studies in small numbers of animals, to longer-term studies of 14 days to three months; repeated dose studies accounted for 75% of dog use when Gad published his guide in 2006. Only when long-term use of the medicine is indicated in humans are toxicity studies of more than six months performed. The need for single-dose, or acute toxicity, studies has since been replaced, meaning that repeated dose studies account for most studies conducted (Robinson et al., 2008). Phillips et al. (2004) provided the range of dose-group sizes used by European companies, as part of a project designed to minimise dog use and harmonise study protocols at an industry level, finding that most studies use four dogs, with the use of two dogs becoming more common.

The "International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use" (ICH) requires that a rodent and a non-rodent species are used for preclinical testing of new medicines (ICH, 1997). The

Table 1.1: Adapted from Gad (2006): Analysis of dog use in drug development projects: data provided by pharmaceutical companies

Type of test	**%**
Safety pharmacology	4.6
MTD/DRF studies	8.4
Repeat dose studies:	
14 days-1 month	28.0
3 months	4.5
6 months	24.1
9-12 months	18.1
Follow-up investigations	11.0

dog is the most commonly used non-rodent species and is therefore considered the "default non-rodent" (Smith et al., 2002), and has been since the 1950s and 1960s (Box & Spielmann, 2005; Gad, 2006). Indeed, Gad (2006) reports their use in medical testing dating back to the 17th Century, the availability of historical data being one of prevailing reasons for dogs' continued use. Phillips et al. (2004) states that the choice of non-rodent species should be based on scientific and ethical considerations, however the dog is commonly used because of its size, temperament and the volume of backdata available (Box & Spielmann, 2005). Table 1.2, adapted from Gad (2006), Box and Spielmann (2005) and Slaughter et al. (2002) highlights some of the common advantages and disadvantages of the dog model. It should be noted that many of these related to the ease of conducting procedures and not to the quality of data or concordance with human data.

Selective breeding and a shared evolutionary environment have resulted in the dog being an invaluable model for human disease and toxicity, due to similar influences acting on genetic evolution. A study commissioned by the International Life Sciences Institute (Olson et al., 2000) found that dog studies were considerably more predictive of human toxicity than rodent studies: a concordance rate with clinical trials of 71% was found for combined rodent and non-rodent studies, with non-rodent studies alone being predictive for 63% of human toxicity. The need to Refine dose ranges so that the necessary clinical signs of toxicity can be identified without causing undue suffering has been identified, with guidance produced by NC3Rs on dose selection (Robinson et al., 2009). A similar concern with body weight loss was addressed in a study by Chapman et al. (2013) with Refined limits being applied to loss in body weight (less body weight loss permitted) without loss of scientific information relating to toxicity.

Sequencing of the canine genome by the National Human Genome Research Institute (Starkey, Scase, Mellersh & Murphy, 2005) suggests that the dog is an "unrivalled model for the study of human disease" (pp. 112) and that dogs share 220 homologous

TABLE 1.2: Advantages and disadvantages of the dog model in safety assessment, adapted from Slaughter et al. (2002), Box and Spielmann (2005) and Gad (2006).

Advantages	**Explanation**
Medium size	Ease of handling, observation and measurement, particularly of haemodynamics
Moderate length of hair coat	Ease of physical examination and procedures
Even temperament and friendly disposition (easy to handle)	Easy to work with (e.g. dosing, blood collection, ECG)
Adaptability to living in groups	Absence of intra-group stress, ease of housing
Satellite animals not needed for serial blood collection	Contributes to Reduction of animal numbers
Volume of back data	Considerable back data available for comparison, contributes to experimental design and Reduction
Selective breeding and shared evolutionary history with man	Increased validity of model
Ease of vomiting	Prediction of emesis in clinical trials
Disadvantages	**Explanation**
Variation in size and body weight	Can result in variations in compound needed and in physiological measurements
Loud, penetrating bark	Requirement for PPE for staff to prevent hearing loss
Cost of acquisition and maintenance	Space and environment maintenance costs greater than smaller species
Greater test compound requirements than smaller species	Greater cost for test compound manufacture
Availability	Bred in smaller numbers than many rodent species, longer gestation
Exercise and housing requirements	Requires considerable space
Societal concerns	The use of dogs may be subject to greater public concern due to their status as a companion animals

TABLE 1.3: Purposes of the experiments involving dogs in EU in 2008, adapted from Pellegatti (2013)

Type	**Number (%)**
Biological studies of a fundamental nature	1814 (8.5)
R&D products and devices for human and veterinary medicine and dentistry	4405 (20.7)
Production and quality control of products and devices for human medicine and dentistry	157 (0.7)
Production and quality control of products and devices for veterinary medicine	207 (9.7)
Diagnosis of disease	1111 (5.2)
Educational and training	362 (1.7)
Toxicological and other safety evaluation	11,077 (52.0)
Others	316 (1.5)
Total	21,312

hereditary diseases with uniform genetic mutations (Zurlo et al., 2011). In particular, dogs have been established as sensitive models for cardiovascular and nervous system changes (Moscardo et al., 2009). Welfare concerns have meant that nonhuman primate use is decreasing, and as such, the dog is becoming the preferred non-rodent model. Dogs are also used for agricultural and industrial chemical testing, veterinary medicine development and testing of medical devices and pet food products (Zurlo et al., 2011), which although investigating different parameters from non-clinical safety testing, still often require that the subjects are a healthy, valid model of the human or dog target organism.

1.1.5 The use of dogs in toxicology

In total, around 12 million animals are used in the EU each year, and 17 million in the US (Taylor et al., 2008) ; dogs constitute only a small proportion of the total animals used. Of all dogs used in studies conducted during the drug development process in the EU, approximately 50% are used for regulatory safety assessment (Table 1.3), a smaller percentage than in the UK. Regulatory safety assessment refers to the range of legal requirements for assessing safety and efficacy prior to human exposure and is governed by the International Convention on Harmonisation (ICH) guidelines. Furthermore, the current protocol for safety assessment of new medicines as set out in the ICH guidelines has changed little over the past decades, in part due to the process required to validate new protocols (Pellegatti, 2013).

Non-clinical development covers all studies 'before man', although animal studies may be conducted along-side clinical studies, with the aim being to explore the mechanism of action, potential toxicity and pharmacokinetics, known as ADME: absorption, distribution, metabolism, and excretion (Pellegatti, 2013). The development of new medicines destined for human use evolves from discovery, through several levels of safety and efficacy testing, before being determined suitable for testing in human clinical trials. The first *in vivo* trials will usually be a small, pilot study in a rodent species (and most often mice), followed by a toxicity study of one- or three-month duration. Longer-term studies are indicated where the medication is considered for long-term use in humans. Progression from the rodent to non-rodent models occurs if the observed side effects are deemed acceptable for the level of benefit obtained from the compound's use. ICH guidelines "Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals" (International Conference on Harmonization, 2008) are the standard by which testing is conducted. The various levels of testing required are laid out in Table 1.4.

Safety assessment is not restricted to the development of new medicines; the safety assessment of pesticides is also a common use of dogs in Europe. Box and Spielmann (2005) report that at least 9,000 animals are needed for each pesticide developed, at least 75% of which represents investigation on reproduction and development. Dogs represent around 0.09% of dogs currently used in Great Britain, potentially resulting in an additional 8 dogs used for each pesticide tested. Within the EU, the REACH (Registration, Evaluation, Authorisation and Restriction of Chemical Substances) legislation will lead to an increase in the safety assessment of pesticides and other chemicals and as such dog use is set to increase in coming years; Pellegatti (2013) estimates that up to 17.6 million animals may be needed to fill this knowledge gap. The use of dogs in the safety assessment of medicines is also set to increase due to a change in legislation requiring "juvenile toxicity" data, safety testing in young animals for medicines which may be used in juvenile humans (Pellegatti, 2013). This potential increase in dog use, in addition to the current global use, suggests that it is critically important to address the current knowledge gap regarding welfare needs and assessment, and the impact of welfare on quality of data output. A clear understanding of the factors which can potentially reduce welfare in the dog is needed along with the promotion of better welfare, before addressing the impact on data output quality.

Table 1.4: Types of dog study adapted from International Conference on Harmonization (2009) and Gad (2006)

Safety Pharmacology	Performed during early development and before a first dose in humans *In vitro* cardiovascular assessment using dog tissues: Purkinje fibre/Langendorff preparation for action potential duration/QT Interval evaluation) Anaesthetised non-recovery studies to assess haemodynamics, electrocardiogram (ECG) and respiratory and renal parameters Telemetry in surgically prepared, conscious dogs to assess cardiovascular system and ECG.
Maximum Tolerated Dose (MTD) **Dose Range Finding (DRF) studies**	Allows selection of dose levels for regulatory studies Target organ toxicity
Repeated dose studies	Durations of 14 days-1 month, 3-6 months, and 9-12 months in general
Juvenile toxicity	A recent requirement for medicines which may be used in paediatric treatments. Uses pre-weaned dogs
Investigational studies	Project-specific Investigating any aspects of toxicity raised by other studies More commonly performed in rodent species but may be needed in the dog
Discovery support	Generation of safety data in early drug development More commonly performed in rodent species but may be needed in the dog

1.2 The dog: Its natural history

All breeds of domestic dog *(Canis lupus familiaris)* are descended from a now-extinct subspecies of the grey wolf *(Canis lupus*, Serpell, 1995). DNA evidence has ruled out any involvement from non-wolf species in the dog lineage (Vilà et al., 1997), although it is unclear precisely which subspecies dogs descended from. Despite historical use of the wolf as a as a model for dog behaviour, the wolf is considerably different from the domestic dog, due to differences in selective pressure in the intervening millennia. It is hypothesized that a genetic bottleneck in the wolf population means that the ancestral sub-population from which dogs are descended would have been more genetically

diverse and may have born little resemblance to the modern grey wolf population (Freedman et al., 2014), meaning that modern wolves are unlikely to provide a valid model of domestic dog behaviour. Therefore comparative studies between grey wolves and domestic dogs are of little use in gaining an insight into the natural behaviour and needs of the dog. Moreover, fossil records indicate that there have been domestic dogs in Europe for around 15,000 years, with dogs last sharing a common ancestor 18,800 years ago, although some dogs last shared a common ancestor 32,100 years ago pointing to multiple domestications in multiple locations (Thalmann et al., 2013). A common physical characteristic of domesticated dogs is paedeomorphism (Gould, 1994), caused by our selective breeding of child-like qualities such as large eyes, preference for eye contact and retaining puppy-like features normally lost in adolescent canids. Selective breeding has also favoured traits such as cooperation and tolerance for living in close contact with humans and other dogs, rarely seen in other, non-domesticated, canids.

While for other laboratory-housed species, in particular primates, we have a wild counterpart from which to obtain information on 'natural behaviour', this doesn't exist for the domestic dog. Instead, we need to piece together literature and information on the natural history of the dog, the process of domestication which has influenced much of its behaviour and physiology, and the behaviour of free-ranging dogs who may be the closest analogue to a wild counterpart. These dogs typically form fluid groups, composed of more stable breeding pairs (Spotte, 2012), in contrast to wolf societies which are formed around a 'family' structure, with one actively breeding pair and previous generations of offspring (Mech & Boitani, 2003). Dogs are also crepuscular, with activity at its greatest around dawn and dusk (Beck, 1973, see also Appendix C for confirmation of this pattern in the laboratory-housed dog). Studies of dog personality traits have failed to establish a 'dominance' trait, a common misattribution resulting from observations of captive-housed wolves (Mech & Boitani, 2003).
'Dominance' is commonly used to explain aggressive or territorial behaviours such as resource guarding. Increased aggression is associated with increased environmental stress and limited availability of highly-valued resources (DeVries, Glasper & Detillion, 2003) so care must be taken to identify the cause of these behaviours, rather than misattributing to 'dominance'. Comprehensive studies of dog personality have identified five major personality traits: "Playfulness," "Curiosity/Fearlessness," "Chase-proneness," "Sociability," and "Aggressiveness". These personality dimensions were developed by Svartberg and Forkman (2002) based on behavioural data from 15,000 dogs of 164 breeds. Although not explicitly tested in this thesis, dimensions of personality such as the curiosity/fear dimension are known to influence an animal's interactions with its environment and may affect its welfare (Panksepp, 2011).

1.2.1 Behavioural development

Puppies' development was first described in phases by (Scott & Fuller, 1974), who identified four phases of development:

1. 0-5 days. Characterised by a lack of response of the audio-visual system.

2. 5-18 days. Adult postural reflexes appear, the eyes begin to open and a weak startle response may be present.

3. 18-28 days. Emergence of positive orientation towards audio and visual responses, puppy reflexes of rooting and urination on stimulation disappear, adult sensory reaction (startle, avoid cliff edge) appear and motor activities emerge.

4. 28 days-adulthood. Adult behaviours emerge. Puppies are capable of recognising humans/littermates visually and auditorily.

Building upon this, Jones (2007) described the stages of development in more detail, detailing the emergence of key reflexes and the responses associated with them (Table 1.5). Introducing gentle stimuli such as changes in sound, texture and light as the eyes and ears open may help to increase resilience in later life by providing a variety of sensory experiences (see Chapter 2).

The emergence of sexual maturity may vary between breeds of dogs, and in the beagle, the onset of puberty occurs between 6-12 months, with sexual maturity reached at between 9-12 months. Growth and food consumption increases from five to approximately ten months, where it begins to level off and most dogs are considered adult at 12 months (Gad, 2006). However, typical use of dogs in safety assessment occurs at an age of nine months or younger. This means that many of the opportunities to influence behaviour and resilience in later life occur at an age when many dogs will be in the breeding facility before being transferred to a dog facility. The stressors associated with transport and acclimatisation to a new facility mean that early opportunities for desensitisation in the breeding facility are particularly important to future welfare.

1.2.2 The dog-human relationship

Humans are an unavoidable component of the laboratory environment, responsible for cleaning and feeding, regulated procedures as well as designated periods of interaction.

Table 1.5: Age at which conditioned responses to stimulation of a sensory system have been elicited (adapted from Jones, 2007)

Sensory system	**Response type**	**Response**	**Day**	**System maturity** (day)
Tactile	Approach	Move/orient towards food	1	1
	Withdraw	Avoid puff of air	10	
Thermal	Approach	Move/orient towards food	7	1
	Defensive	Defensive reaction	17	
Vestibular	Approach	Move/orient towards food	10	1
Taste	Approach	Sucking response	1	1
	Withdrawal	Move away	14	Undetermined
Olfactory	Approach	Move/orient towards food	1	
	Withdrawal	Move away	13	
Vision	Approach	Move/orient towards food	21	49-56 days
	Withdrawal	Move away	21	Eyes open 7-19 days
	Defensive	Struggle	27	
Audition	Approach	Head orientation	21	20 days
	Withdrawal	Move away	18	Ears open: 12-14 days
	Withdrawal	Withdraw leg	27	

Humans and domestic dogs have been closely associated for 12-15,000 years (Serpell, 1995), and during this time, dogs and humans have developed a close and often mutually-beneficial relationship, for example cooperation in hunting or herding of other animals (Bradshaw, 2011). This has resulted in significant adaptation, with selective breeding for traits promoting close cooperation with humans. The result of this history of cooperation in a shared environment has led to the dog developing abilities to communicate with and understand humans which in some ways exceed those of nonhuman primates. Dogs demonstrate an understanding of human eye gaze similar to that of human infants and exceeding that of the dog's ancestor, the wolf (*Canis lupus*) or the chimpanzee (*Pan troglodytes*, Hare & Tomasello, 2005; Kubinyi, Viranyi & Miklosi, 2007). Dogs also demonstrate patterns of behaviour similar to those of attachment in human infants in tests of Ainsworth's 'Strange Situation' paradigm (Gacsi et al., 2005; Palmer & Custance, 2008), using a human caregiver as a secure base to explore a novel environment and showing distress at separation. This close relationship is potentially a critical welfare problem in the laboratory environment where dogs are subject to adverse interactions with staff. While laboratory-housed dogs do not have the same relationship with a human carer as a pet dog might, interactions with humans still have the potential to be a tool for improving welfare. However it also provides an opportunity to increase welfare in a manner not possible

with species which view humans as a threat; an improved relationship between dog and handler leads to the use of the handler as a 'secure base' during adverse events.

1.2.2.1 The effects of human interaction on dog welfare

Table 1.6 summarises the findings of studies investigating the effects of human interaction programmes on behaviour and physiological measures in pet, shelter, working and laboratory-housed dogs. Studies differed in the frequency, duration and content of human interaction periods, and in the results found. Studies investigating the effect of a single human interaction period have found an immediate positive change in physiological and behavioural signs of stress, however no lasting effects were found. This conclusion was also drawn by Taylor and Mills (2007a) in a review of human interaction programmes. In contrast, Normando et al. (2009) found that the changes in behaviour exhibited by dogs which underwent the human interaction programme were maintained for two weeks following termination of the programme. Importantly, Normando et al. (2005) found that interruption of a regular walking programme resulted in an increase in cortisol levels in laboratory-housed beagles, so once started, it is important that a programme is maintained. However, all of these studies found a positive impact from brief interventions, suggesting that such interventions should be beneficial in the laboratory environment.

TABLE 1.6: Summary of the effects of human contact and interaction on dogs in various environments

Population	*n* Dogs	Frequency of interaction(s)	Procedure	Measures	Findings	Study
Laboratory dogs (beagles)	Control = 12, conspecific interaction = 12, intensive handling = 12, chew toys = 12	30 sec per day for 2 months	30 sec of grooming and handling per day	Behavioural observations, technician ratings, salivary cortisol	Behavioural and time budgets did not vary, except in the chew toy group. Handled dogs scored more highly on friendliness, approachability and technicians reported interactions to be rewarding. No differences were found in salivary cortisol	Hubrecht (1993)
Shelter dogs (multiple breeds)	Training group = 20, control group = 20	5 days per week for 8 weeks	20 minutes of positive reinforcement per day	Observation during 4 tests: novel environment, novel person, novel object and following an air horn	Dogs which underwent training showed fewer non-directed licks and escape attempts in novel person test	Hennessy, Voith, Young et al. (2002)
Shelter dogs (multiple breeds)	Human interaction = 9, control = 13	Once weekly for 6 weeks	One 15 minute interaction session per week	Behaviour, time spent out of sight	Human interaction group had an increase in tail wagging, inactivity and a decrease in time spent out of sight for up to 2 weeks after study	Normando et al. (2009)
Shelter dogs (multiple breeds)	Human petting = 44, control = 30	Once	Two blood samples taken 20 minutes apart, with 20 minutes either in home kennel or of human petting between samples	Blood cortisol levels	Increase in blood cortisol levels during second sample following time in home kennel but no increase following human petting	Hennessy, Williams, Miller, Douglas and Voith (1998)
Shelter dogs (multiple breeds)	Human interaction group = 68, control = 62	Once	45 minutes of human interaction on second day in shelter	Salivary cortisol	Human interaction group showed a decrease in salivary cortisol from day 2 to day 3. Control group showed an increase between days 2 and 3	Coppola, Grandin and Enns (2006)
Shelter dogs (multiple breeds)	Human interaction = 9, Control =8	Once a week for three weeks	15 minutes of human interaction once a week	Behavioural observations, responses to novel stimuli, rehoming rating	HI group had a decrease in rate of stereotypies, agitated locomotion and increase in tail wagging. HI dogs had an increase in rehoming score.	Hall (2010)
Shelter dogs (multiple breeds)	HI = 10, control = 10	3 days a week for 8 weeks	25 minutes including play, grooming and walking	Heart rate variability (HRV), salivary cortisol and behavioural testing	HI group showed increased sociability, some reduction in HRV during testing and HI group showed significant decreases in salivary cortisol levels.	Bergamasco et al. (2010)
Pet dogs (multiple breeds)	18 dogs	Once	30 minutes of affiliative behaviour between humans and dogs	Blood pressure, β-endorphin, oxytocin, prolactin, phenyl acetic acid, dopamine and cortisol levels.	For dogs: decrease in blood pressure, increase in β-endorphin, oxytocin, prolactin, phenyl acetic acid and dopamine levels.	Odendaal and Meintjes (2003)
Service dogs (German Shepherd Dogs)	Police dogs = 53, border guard dogs = 31	Once	For one 3 minute period, handlers and dogs played with tug toy	Salivary cortisol before and after play session, behavioural dimensions for dogs and handlers	Border guard dogs: handlers exhibited more petting and enthusiasm, dogs' salivary cortisol decreased. Police dogs: Handlers exhibited more discipline and commands, dogs' salivary cortisol increased.	Horváth, Dòka and Miklòsi (2008)

The quality and quantity of an intervention clearly has an influence on its efficacy. Hennessy, Voith, Young et al. (2002) found only two behavioural differences following intensive positive-reinforcement training when compared to a control group which didn't receive training (fewer non-directed licks and fewer escape attempts), while Normando et al. (2009) found differences in behaviour (increased tail wagging and increased inactivity) and proximity (less time spent out of sight) following fewer interaction periods, but which involved positive social contact. Differences in the behaviour of a human handler (when measured in dimensions of enthusiasm and discipline) can have an effect on whether a play session increases or decreases a physiological measure of stress, and training alone may not be sufficiently positive to improve welfare (Horváth et al., 2008).

Wells (2004b) suggested that play especially may be a particularly helpful tool in interaction with dogs in shelters, as it establishes "appropriate dog-human relationships" and prepares a dog for rehoming. In the laboratory environment, a positive social bond between dogs and handlers may facilitate learning in training for procedures, in completing husbandry protocols without introducing an element of stress through fear responses, and in providing a tool to increase resilience when welfare may be compromised by aversive events such as conspecific isolation. Dogs housed in laboratories and rescue shelters may be provided with little opportunity to interact with humans (from between 0.3 and 2.5% of time observed, Hubrecht, Serpell & Poole, 1992). Human contact clearly has the ability to mitigate the effects of some adverse events upon welfare, particularly through the use of a human carer as a 'secure base' and to provide positive experiences in the laboratory environment. The ability of human contact to cause positive changes in haemodynamics and endocrine responses to stress suggests that the structured implementation of a Refinement protocol may improve welfare and data quality in a feasible manner. However care must be taken when developing such a Refinement protocol so that it is of a suitable frequency and duration and that it is feasible, since the loss of predictable positive human interaction may be more detrimental to dog welfare than never having it at all.

1.2.3 The beagle

The beagle possesses many of the characteristics which have been cited as advantages of using the dog in scientific research (Table 1.2). Its small size and amiable temperament make regular handling and procedures easier to conduct than in large dogs, while its history as a pack-living, working dog makes housing in large groups possible. The history of the breed can be traced back to the long-extinct Talbot dog (c. 11th Century) which gave rise to the Southern Hound. The beagle as a breed can

first be identified from around 1830 (Youatt, 1852). The continued popularity of the breed, both as a pet and as a working dog, can be attributed to its ability in scentwork and its temperament.

Beagles generally have a highly excitable temperament, scoring highest in a test of 'excitability' (Fogel, 1990), potentially resulting in physiological markers of excitement acting as confounding variables in the laboratory environment. Conversely, beagles are also prone to displaying "agreeableness" as a personality dimension, and are not prone to aggression (Kraeuter, 2001), making them ideal for an environment which can involve intensive handling and group living. Beagles have a strong scenting drive: in a study reported in Scott (1965), beagles were able to find a mouse in a one-acre field in under a minute, compared to the next fastest breed, fox terriers which took 15 minutes. High drive to perform natural behaviours and the energy levels associated with a working, scenting breed should be taken into account in any programme of Refinements, providing opportunities to exercise and forage.

1.3 The welfare of dogs in laboratories

The study of animal welfare covers a broad range of concepts, and a single approach has yet to be agreed upon (Fraser, 2008). This section discusses an approach to animal welfare based on the ability to express natural behaviour, how the animal feels and physical health. The foundation of modern approaches to experimental animal use, the 3Rs, are based on a societal concern for animal welfare, an idea that we should not subject animals to potentially painful procedures when there is an alternative, and to reduce the numbers and severity of these procedures (Russell & Burch, 1959). This concern for animal welfare comes from an enduring concept of animal feelings, a similarity to our own subjective experience of positive and negative emotions, rather than a difference in the 'lower' animals (Paul, Harding & Mendl, 2005).

Welfare can be understood in terms of physical health, and in terms of subjective experience. Dawkins states that it is not possible to have good welfare while experiencing poor physical health, yet physical suffering and emotional suffering are not synonymous (for example, physical health can be poor but analgesia prevents emotional suffering and therefore poor welfare, Dawkins, 2008a).

1.3.1 Animal welfare definitions and assessment

'Welfare' is a term which may have various, colloquial, uses in society but in order for it to have scientific validity and to allow comparison of welfare across situations, it

must have an objective definition. Indeed, there is still some debate over the definition and assessment of welfare in animals (e.g. Fraser, 2008; Hawkins et al., 2011). Broom (1986) described welfare as a term used to describe an individual's state in relation to its attempts to cope with a situation. Welfare does not reflect the external circumstances *per se* but rather how effectively an individual is coping with them and the resulting impact on fitness (defined in an evolutionary sense). Therefore, welfare cannot be 'given', rather it is a characteristic or mental state of an individual and it is accepted that it is not possible to give 'better' welfare, only to affect external circumstances in such a way as to promote positive welfare (Broom & Kirkden, 2004). It is also well accepted that 'welfare' also describes a continuum from negative to positive, rather than a desirable condition (for example, the common confusion between 'animal rights' and 'animal welfare'), and that therefore external circumstances and internal coping mechanisms interact to produce a welfare state somewhere on this continuum. Welfare can be considered to be compromised when an individual experiences physical injury or illness, cannot access adequate food or housing, or experiences stress with which it cannot cope (Broom & Kirkden, 2004). Welfare in animals is used in the same sense as in humans and is analogous to terms such as 'quality of life' and 'well-being' (Christiansen & Forkman, 2007; Taylor & Mills, 2007b).

A valid system of measuring animal welfare should be free of moral or ethical considerations (Broom & Kirkden, 2004) and instead concentrate on empirical evidence. Each system of measurement for welfare has the potential to make important contributions to the study of welfare, however none can be used in isolation and each has its shortcomings. The origin of behaviours (the function of the behaviour to the animal) as well as a quantitiative measurement is necessary in an assessment of welfare. For this reason, the Framework developed in this thesis incorporates various aspects of behaviour, emotion and physical health to understand the nature of welfare in the dog. Latham (2010) described a 'toolbox' of welfare measures: indirect measures of welfare, including behaviour, physiological, neurophysiological and chemical indicators all of which can be used to answer Dawkins' questions of "is the animal healthy?" and "does the animal have what it wants?" (Dawkins, 2004, 2008b).

Distress can be defined as the result of exposure to stressors in the environment which over-taxes an individual's coping systems and reduces, or seems likely to reduce, its fitness (Broom, 2006). "Suffering", distinct from distress, is the internal emotional state which results when an animal is unable to cope with challenges in the environment which chronically overwhelm coping mechanisms. Suffering is an unpleasant subjective state which occurs when an animal is unable to carry out the actions which would otherwise improve its fitness or environment (Dawkins, 2008a).

Although understanding the subjective state of nonverbal animals is difficult, there are methods of assessing welfare through physiological and behavioural indicators of pleasure and stress, physical health, physical fitness, assessment of preferences, ability to express normal functions and attempts to cope (e.g. Broom, 1991a; Broom & Kirkden, 2004).

Welfare state can be inferrerd through direct measurement of activation of the hypothalamic-pituitary-adrenal axis (HPA axis) by measuring levels of adrenalin, glucocortisoids and adrenocorticotropic hormone (ACTH, Broom, 1991b) and the effects of prolonged exposure to these hormones such as immunosuppression, reduced growth and reproductive success and physical exhaustion (Dawkins, 2013).

1.3.1.1 Emotion

Broom (1998) stated that feelings are a biological process which have evolved to improve fitness; feelings are used to make decisions which have long-term implications, physical feedback is used for short-term decisions (Broom & Kirkden, 2004). In laboratory-housed animals, feelings reflect experiences in the environment and the strategy adopted for coping (see Chapter 4 for discussion of this). While animal 'consciousness' remains a contentious term, there is debate about whether consciousness is even relevant to the discussion of welfare (Dawkins, 2006). Russell and Burch (1959) believed that as animals are less able to understand the world than us, that they are at greater risk of poor welfare and of failing to adapt or cope:

"The sequence of moods in a lower animal, however, is rigidly controlled by internal and external changes according to a code of rules, largely preset for a given species." pg. 17

Although there is discussion about whether or not nonhuman animals can suffer if we have no evidence that they are capable of experiencing *conscious* emotions, there is sufficient evidence that they possess the same brain structures necessary for emotion (Panksepp, 2011). In the dog in particular, fMRI studies have shown activation of the brain in response to positive and negative stimuli in a similar way to humans. Andics et al. (2014) investigated responses to vocal and nonvocal stimuli. Not only were human vocal stimuli found to be processed in the temporal pole, in the anterior temporal lobe, but in both humans and dogs, the primary auditory cortex was activated when listening to emotive sounds. Similarly, work by Berns, Brooks and Spivak (2012) found that both presentation of a positive visual signal and a familiar human scent activated the caudate nucleus, the same structure which in humans,

activates in response to rewarding stimuli. Although animals are non-verbal, and verbal communication is the "gold standard" of measuring emotion (Paul et al., 2005), to discount the similar experience of animals unable to communicate verbally would be negligent and unethical.

The dog shows striking similarities to humans in behavioural responses to stimuli which elicit positive and negative affect. Patterns of secure and insecure attachment are seen in dogs in relation to their human owners, and distress exhibited when left alone in an unfamiliar environment (Gacsi et al., 2005). Extensive investigation of hemispheric lateralisation in dogs has demonstrated a pattern that is seen across many species and much like emotion relates to the interpretation of stimuli. Emotive stimuli are processed by the right hemisphere in an approach/withdrawal decision-making paradigm, whereas non-emotive or cognitively challenging stimuli are processed by the left hemisphere. This phenomenon has been found in visual (e.g. Siniscalchi, Sasso, Pepe, Vallortigara & Quaranta, 2010), auditory (e.g. Siniscalchi, Quaranta & Rogers, 2008), and olfactory processing (e.g. Siniscalchi et al., 2011) as well as paw preference (e.g. Branson & Rogers, 2006). Rogers (2010) has suggested that lateralised behaviour can be used as a welfare indicator. This common processing of emotive stimuli suggests that the architecture underlying emotion is similar across many species. Links have also been found between laterality and personality dimensions in guide dogs (e.g. Batt, Batt, Baguley & McGreevy, 2008b) as well as success rates in training. Taken together, the lateralised processing of emotive stimuli in dogs suggests a common emotional processing system with humans (e.g. Canli, Desmond, Zhao, Glover & Gabrieli, 1998) and other animals (e.g. Rogers & Vallortigara, 2008; Rogers, 2008)

1.3.2 Using behaviour to assess welfare based upon feelings

Behaviour is the most established and perhaps the most useful measure of welfare. Behaviour can be observed without specialist or invasive equipment, can provide instantaneous information on an animal's reaction to stimuli and when observed using an agreed coding scheme, is free from subjective bias. Conversely, one of the drawbacks of using behaviour is that it tells us little about the internal state of the animal. Behaviourist approaches state that we can only observe behaviour without inferring meaning. This stance is still adopted by those measuring behaviour for welfare, as in isolation, we cannot know what a particular behaviour means to an animal, only through observation of other factors can we learn to associate behaviours with internal states. The ability to express natural behaviour is accepted as one of the foundations of good welfare (Poole, 1992). Dogs have evolved to express a pattern of behaviours necessary to ensure their fitness and the inability to do so results in

frustration and the loss of homeostasis (Fraser, 2008). Although the function of such behaviours may have been lost in the process of domestication, they should be considered no less important to the animal; the ability to express natural behaviour provides the animal with a range of coping strategies in the presence of a stressor and prevents the frustration associated with being unable to do so (Koolhaas et al., 1999). The laboratory environment, by its nature, restricts the ability of dogs to express the full range of normal behaviours, and so seeking to promote the range of behaviours seen in free-ranging, pet or working dogs is neither possible or strictly desirable. A range of adverse events may be experienced by these dogs, for example predation or aggression from conspecifics, and mimicking these in the laboratory is not desirable for welfare, however allowing dogs the opportunity to use their natural behavioural repertoire to cope with stressors *is*.

The literature identifies aspects of the laboratory environment which have the potential to affect the welfare of many laboratory-housed species, and in particular the dog. Table 1.7 outlines some typical components of safety assessment studies, highlighting the elements which have the potential to cause a change in welfare. Unlike some other laboratory-housed species (e.g. nonhuman primates, rodents), there is a paucity of data about how dogs respond to specific stimuli or their environment. What we do know about housing (Hubrecht et al., 1992), socialisation with human carers (Hubrecht, 1995a) and responses to aversive stimuli (Beerda, Schilder, Van Hooff, De Vries & Mol, 1998) is that factors to relating spatial and social restriction, unpredictable aversive events and restraint are perceived as stressful by dogs and have the potential to reduce welfare. Human contact (Hubrecht, 1993), conspecific contact (Hubrecht et al., 1992) and environmental enrichment (Schipper, Vinke, Schilder & Spruijt, 2008) can all increase welfare.

TABLE 1.7: Potential welfare impact of study activities, adapted from Gad (2006); Everds et al. (2013)

Activity	Occurrence	Impact
Husbandry		
Home pen cleaning	Daily clean, weekly wash	Disruption of environment, associated handling
Health checks	Conducted weekly or monthly for non-study animals	Handling and restraint may cause distress
Housing		
Group-housing	Typical for non-study animals	Poorly matched groups may lead to conflict with pen mates
Single-housing	For veterinary or study reasons	Social isolation and restriction to smaller pens significant stressor for highly social species like the dog
Home pens	Various home pens used to house non-study and study animals	Minimum size requirements may not meet to exercise and range needs of the dog Noise is a potential welfare concern Pens often consist of hard surfaces, may lack suitable bedding and enrichment
Typical four-week study		
Daily dosing	for 29 or 30 days	IV or oral dosing may cause distress, associated symptoms from compound
Daily observations	Pre-treatment, x2 daily	Potentially unpredictable human presence in unit
Physical examination	Pretreatment, after dosing	Restraint may cause distress, unpredictable relationship between restraint and procedures
ECG	Pretreatment, after dosing on set days	Requires restraint, often in a sling
Opthalmic examination	Pretreatment, set point in final week of study	Requires restraint
Body weight	Pretreatment, weekly, before scheduled sacrifice	Often taken during other health examinations
Feed consumption	Pretreatment and daily during study	Requires single housing to record individual consumption
Clinical pathology	Pre-treatment, set points during study, prior to sacrifice	Blood sampling may cause distress
Urine collection	As above	Requires catheterisation, invasive and potentially distressing; or use of metabolism cages
Sacrifice	Usually staggered across last days of study	Euthanasia is the end point of most safety assessment studies, must be conducted in a manner consistent with 'a good death', typically barbituate overdose followed by exanguination
Dose molecular/ pharmacokinetic studies		
Dosing	Single dose by IV or oral route	IV or oral dosing may cause distress
Pharmacokinetic sample	Blood collected at specified time after dosing	Blood sampling may cause distress
Clinical observations	At set time points after dosing to detect emesis or diarrhoea	Requires single housing
Metabolism	Collection of all urine, faeces and vomit following dose	Requires restriction to metabolism cage (see single-housing)
Safety Pharmacology		
Surgical preparation	Telemetry device implanted before use	Post-surgery recovery may involve pain, distress and single housing
Daily dosing	for duration of study	IV or oral dosing may cause distress
Remote collection of telemetered data	Pretreatment and on set days following dosing for up to 24 hours	May require single housing and removal of pen objects

There are common themes in the literature regarding the behaviours expressed when an animal is restricted in its ability to express natural functions, is unable to satisfy social or physical needs or experiences chronic stessors and distress. For example, vigilant behaviours (being alert to the surroundings, observing stimuli) are common to many species (e.g. Dwyer, 2004; Welp, Rushen, Kramer, Festa-Bianchet & De Passille, 2004) and are more commonly seen in situations where stressors are present (especially unpredictably so), or there is a potential threat to the animal (Rushen, Taylor & de Passillé, 1999). In the laboratory-housed dog this manifests as sitting or standing while alert (Hubrecht et al., 1992), as opposed to relaxing without orientation. While vigilance is clearly a behaviour which can promote survival, it becomes maladaptive where there is no actual threat, but a perceived threat, such as an aversive procedure or presence of staff, removes the ability of the dog to relax. The absolute presence or absence of such a behaviour is not in itself maladaptive or indicative or poor welfare, but heightened vigilance seen throughout the day, in the absence of stimulation indicates that subjectively, the animal perceives a threat.

Conversely, relaxed, but not apathetic, behaviour and an interest in surroundings are commonly seen in situations where there are no threats or social unrest (e.g. Boissy et al., 2007). The provision of an appropriately stimulating environment, for example bedding which can be manipulated and toys which encourage foraging, allow animals the opportunity to exhibit these behaviours, promoting positive emotional states such as happiness, satisfaction and positive anticipation (Poole, 1997; Boissy et al., 2007). In the dog, this manifests as resting behaviour or interacting with the environment, either by manipulating objects such as toys and bedding (not to be confused with agitated behaviours such as repetitive digging) or by air scenting, as olfaction is used by the dog to investigate its surrounding. Interaction with the environment may also occur when a disturbance has occurred in the environment, a response commonly seen in rats following removal of dirty bedding (Balcombe, Barnard & Sandusky, 2004), and so this may be considered a context-specific behaviour (see Chapters 5-7 for examples of this).

Dogs are highly social, forming fluid social groups and stable bonds (Spotte, 2012); they are prone to separation anxiety when prevented from having social contact with conspecfics or human carers (Flannigan & Dodman, 2001). Social isolation (physical, but not always visual or olfactory, separation from conspecifics) resulting from single-housing is recognised as a particularly stressful event in the life of the laboratory dog by the European Directive legislation, although the effects of single-housing are rarely investigated. Hubrecht et al. (1992) found that single-housed dogs were apathetic and had a restricted behavioural repertoire in comparison to group-housed conspecifics; the absence of pen mates removes the ability to express social behaviour and the ability to use others as a social buffer to stress (DeVries et al., 2003). Social

interactions between conspecifics are, therefore, more likely to be amicable than agonistic in the absence of stressors in the environment. In the dog, this is seen as calm interactions which may involve sniffing, grooming or play (Beerda, Schilder, Van Hooff & De Vries, 1997). An increase in tension caused by stressors can result in agonistic interactions becoming more frequent, and is associated with increased cortisol (DeVries et al., 2003). These may involve growling, biting, resource guarding or attack. These behaviours should not be confused with play fighting, which is common in dogs and is not a sign of aggression (Jones, 2007).

Some functional behaviours can become maladaptive when an animal becomes distressed, as is the case with stereotypic behaviours. A stereotypy is a "repeated, relatively invariant sequence of movements which has no obvious function" (Broom & Kennedy, 1993, pg. 151) and is seen almost exclusively in restricted and captive animals, mentally ill or handicapped humans and individuals given stimulant drugs (Hansen & Jeppesen, 2006). A stereotypy may often be a frustrated attempt to complete a natural behaviour pattern in a situation which prevents this, or a repetitive escape attempt (Broom & Kirkden, 2004). However, while the presence of stereotypies indicates poor welfare, stereotypies may themselves improve welfare as a form of self-enrichment or via the "mantra effect", as seen in humans, especially autistic individuals (Mason & Latham, 2004). The frustrated attempt to display a behaviour may in itself become soothing, and a coping mechanism.

There are also physiological correlates to stereotypies, such as positive correlations with corticosteroids (e.g. Hansen & Jeppesen, 2006; Broom & Johnson, 1993), although these are not clear (see 1.3.3 for further explanation). Stereotypies not only reflect frustration but changes in CNS function which persist in the absence of the original stimulus (Garner, 2006). There is evidence that animals which exhibit stereotypic behaviours have lower physiological measures of distress (e.g. Blackwell et al., 2010) as actively coping with the stressor by performing the repetitive behaviour reduces the experience of distress and restores homeostasis. Animals which have no control over aversive stimuli in their environment (as in learned helplessness, see Mason & Latham, 2004) may become apathetic, appearing to be unreactive while maintaining an internal state of distress. The role of stereotypies in measuring welfare is therefore not clear; however what is clear is that stereotypies indicate that in the present, or in the past, the animal has had to develop coping mechanisms to deal with a stressor in the environment. When these behaviours manifest in response to a specific stimuli in the laboratory environment, care should be taken to identify the root cause and develop Refinements to improve welfare. The lifetime experience of animals which have experienced chronic stress must be considered when assessing their current

welfare. A lack of responsiveness to stimuli may not represent a lack of stress, but learned helplessness, and these animals may be more in need of interventions.

When measuring stress and its relevance to welfare the function and effects of a particular change in behaviour for the animal concerned should be understood. For example, while Broom and Johnson (1993) suggested that any experience of stress is indicative of poor welfare, indicating a failure to cope, Wiepkema and Koolhaas (1993) and Moodie and Chamove (1990) have suggested that manageable stress may in some circumstances be positive for the animal because of its adaptive function. Moderate levels of stress may provide a coping mechanism by allowing the animal to engage in an active coping strategy, such as increased activity. Of course, high levels of stress or chronic stress are not adaptive and may result in learned helplessness, immobility and self-mutilation (Fraser & Broom, 1990). Selye (1946) differentiated stress from 'eustress', or good stress, for example the experience of accomplishing a difficult task, for example an environmental enrichment task. Promoting manageable stressful occurrences may increase the ability of the animal to cope with challenges and ultimately be less harmful than attempting to remove all stressful experiences from the environment, as this is unlikely to be possible. The principle of desensitisation is to pair an aversive or stressful stimulus with a pleasant stimulus or reward, and is one method of making subsequent stressful experiences more manageable. Exposure to the stressor does not overwhelm and over repeated exposures, the animal develops an ability to cope.

Beerda, Schilder, Van Hooff et al. (1999) and Beerda, Schilder, Bernadina et al. (1999) investigated the effects of past experiences of stressors by comparing dogs housed outdoors in groups in either good or bad weather which were then moved to individual indoor pens. The good weather group showed greater physiological (cortisol) and behavioural indicators of distress (paw lifting, autogrooming, vocalising, low posture, stereotyping) upon relocation than the bad weather group. This suggests that it was not only the change of social and physical environment which caused distress, but also the perceived loss of valued resources. This also suggests that the dogs which were housed outdoors in poor weather may have developed coping mechanisms which they transferred to the indoor individual housing condition. There is an important point to note here that current welfare may be influenced by previous levels of welfare, and previous experience of stressors; as in this example, a reduced rate of response to an adverse change in circumstances does not necessarily reflect a lack of change of welfare state.

1.3.3 Using physiology to assess welfare based upon feelings

The physiological response to stress, much like the emotional response to stress, can be thought of as an attempt to change to maintain homeostasis. In this way, physiological changes in response to potential stressors can be used to identify changes in welfare. Selye (1946) introduced the concept of the 'general adaptation syndrome', a change in physiological functioning designed to promote survival, and highlighted the role of increased endocrine function in disease models of hypertension and autoimmune disorders.

One of the most commonly used physiological measures of welfare is cortisol, a glucocorticoid released in response to physical or psychological exertion. The hypothalamic-pituitary-adrenal (HPA) axis activates in response to a stressor, increasing endocrine function and blood levels of cortisol which are detectable in blood, urine and saliva (Beerda, Schilder, Janssen & Mol, 1996). Haverbeke, Laporte, Depiereux, Giffroy and Diederich (2008) studied changes in cortisol and behavioural indicators in chronically stressed dogs (a cohort of 27 military dogs), exposed to a startling stimulus, referred to by the authors as a 'challenge'. Challenged dogs demonstrated increased locomotion, circling, nosing, body shaking, yawning and displacement behaviours, while unchallenged dogs demonstrated low posture, autogrooming, corprophagy, vocalising, paw lifting, high levels of locomotion or inactivity, nosing, urinating, and stereotyping. In response to the stressor, dogs stood less, were more active and exhibited more low posture and less tail wagging.

However, as with physical measures of stress such as heart rate (Fallani, Prato Previde & Valsecchi, 2007), changing cortisol levels can be caused not only by distress, but by increased activity (Beerda et al., 1998), positive anticipation (Boissy et al., 2007), and other physiological exertion such as aggression (Rosado et al., 2010) or reproductive stress (Ziegler, Scheffler & Snowdon, 1995). The physiological response to (bad) stress and (good) eustress can be difficult to differentiate. Conversely, chronic states of distress and activation of the HPA axis can lead to a dampening of the cortisol response; so-called 'burnout' is well-documented in humans (Pruessner, Hellhammer & Kirschbaum, 1999), meaning that the absence of a response can be wrongly interpreted as a positive welfare state. As with stereotypic behaviours, raised cortisol can indicate that an animal is actively coping rather than exhibiting learned helplessness (Blackwell et al., 2010). The integration of measures of welfare is necessary to understand their meaning in welfare assessment.

What is clear is that where cortisol is chronically raised, a state which can lead to damage to organ systems, the cause should be identified and addressed to prevent the

health of the animal being affected as well as its emotional state. Barnett and Hemsworth (1990) stated that an in mammals, an increase of more than 40% above baseline levels of cortisol should be considered detrimental to the immune system. Therefore, especially in the case of laboratory dogs which are used as models for human subjects, a substantial increase in cortisol levels, whether as a result of positive or negative arousal, should give cause for concern when considering the suitability of the dogs as models.

Beerda, Schilder, Van Hooff et al. (1999) studied increasing levels of chronic stress, caused by increasing levels of austerity in housing conditions, across four groups. The relationship between measured urinary cortisol levels and activity was found to be complex. Dogs were found to have higher cortisol on days in which they were apathetic, but dogs which were overall more active also had higher levels of cortisol than inactive dogs. This complex relationship between activity and chronic 'burn out' means that cortisol is an unreliable indicator of welfare state. Blackwell et al. (2010) found that increased urinary cortisol was associated with the rapid shaping of behaviour, and learning, in a population of shelter dogs. Animals which failed to learn a simple task over time were highlighted as a cause for concern because of this burn out effect. When we consider that laboratory-housed dogs are asked to adapt to compliance with husbandry and procedural protocols, their ability to learn being reduced by distress may mean that they are less able to predict and control their environment. The use of positive reinforcement training protocols (as used in Chapter 7) would be *ineffective* when the dogs' welfare is so otherwise compromised that they are unable to learn. The subjective experience of poor welfare, manifested in a negative emotional state, causes dogs to become risk averse, reducing their ability to learn. This relationship between chronic exposure to stress and inability to learn becomes apparent later in Chapter 4.

It is worth noting that physiological measures can be used to not only assess welfare, but also the effects of human interaction interventions, which can attenuate the effects of chronic stress as measured through cortisol levels. Hennessy, Voith, Hawke et al. (2002) found that post-exposure to a stressor, shelter dogs' plasma cortisol doubled, while there was no change found for dogs which had undergone a human interaction programme. The human interaction programme appeared to have the function of increasing the dogs' ability to cope with a novel stressor. In the context of the laboratory environment, increasing the ability of dogs to cope with stressors such as an unfamiliar equipment or people must be considered an important intervention.

1.3.4 Developing a framework to harmonise measures of dog welfare in the laboratory

There are ample guidelines suggesting that factors relating to the housing, husbandry and handling of laboratory-housed dogs be promoted to ensure good welfare (e.g. Prescott et al., 2004), however these are often based on anecdotal evidence of efficacy, from practitioners, and expert opinions, with little empirical evidence of their efficacy, uptake across industry has been variable and legislative guidelines often do not mandate specific Refinements. This lack of compelling evidence also means that the standard of harmonisation between countries is poor. For example, under A(SP)A (1986), a minimum floor size of 2.25m^2 per dog (4.5m^2 when singly-housed) is mandated (1.5m^2 for post-weaned breeding stock). Conversely, the minimum pen size for a dog in the USA (under US AWA Code of Federal Regulations Section 96FR3.6) is "(dog length + 6") x (dog length + 6") /144", equivalent to 6.25 inches^2 for a dog of 24 inches length, considerably smaller than is allowed under UK or European legislation. Restriction of home pen or exercise space is detrimental to welfare in the dog, with free-ranging dogs having a home range of many hectares and travelling up to 30km in a single foray (Meek, 1999). Hubrecht et al. (1992) stated that there is no experimental basis for the minimum housing requirements set out in UK legislation. However, their study comparing group-housed (5 dogs per large pen) and single-housed (one dog per small pen) laboratory dogs found that the single-housed dogs were more inactive and had a more limited behavioural repertoire than their group-housed conspecifics. Contrasting information is found in guidelines for best practice. While Prescott et al. (2004) recommend the use of positive reinforcement protocols for habituating or desensitising animals to dosing protocols, Gad (2006), in a comprehensive manual on laboratory animal science, recommends oral dosing "by way of a rubber bit placed between the teeth, or towels" (pg. 577). This is a use of force and lack of Refinement which is incompatible with modern practices and yet comes from a more recently-published manual, although written by an author from the USA. With allelomimetic barking acknowledged as potential welfare concern (Sales, Hubrecht, Peyvandi, Milligan & Shield, 1997), facilities in Europe are adapting housing with increased visibility (as pictured in Chapter 3), while Gad (2006) recommends that barking can be controlled by partial ventriculocordectomy without affecting "well-being". Reliable sources of information on best practice are available, for example the NC3Rs in the UK, however without a demonstrable link between good welfare and improved quality of data output, many of these potentially important Refinements will be lost in a literature body based on anecdotal evidence.

What is apparent from reviewing the literature is that there is no single, gold

standard, measure of welfare. Using Broom's definition of welfare, *"the ability of an animal to cope with its environment, and the corresponding effects on its fitness"* (pg. 147), it is clear that in the context of the laboratory-housed dog, there is a pressing need to understand how dogs cope and what can be done when they fail to. There is less agreement about how to measure welfare. Behaviour is subject to criticism where anthropomorphism interferes with our ability to understand the meaning of a behaviour for an individual. Emotion, although no longer so controversial in animals, is not easy to measure or interpret. Physiological measures such as heart rate or cortisol can also be ambiguous due to the similar responses to stress and eustress, as well as in response to increased physical activity. However, chronic activation of the HPA-axis which results in an inability to return to "baseline" in an animal model should be of real concern. Our perceived baseline is changed, although unknowingly. The sensitivity of measures will be affected, as will our ability to interpret results. It is still vitally important to develop reliable, unbiased welfare assessment tools in order to understand how welfare impacts the physical health of animal models and our ability to conduct "good science" by taking account of all variables.

The approach taken by this project is to quantify easily-identifiable measures of behaviour which are suggested *a priori* by the literature to indicate positively- or negatively-valenced welfare and associate these with established measures of emotional state, while understanding the relationship between these measures and those which affect the physical health and data output of the dog, in particular cardiovascular function. A Welfare Assessment Framework integrating these factors can be used by welfare scientists, technicians and care staff to monitor and identify animals at risk of poor welfare, while ensuring the best possible quality of data is obtained from their use. By reducing unwanted variation caused by poor welfare, Refinements can be implemented which may lead to a Reduction in dog numbers. The following chapter describes in more detail the relationship between welfare and data output.

CHAPTER 2

The ethical and scientific importance of dog welfare in laboratories

"The central problem, then, is that of determining what is and what is not humane, and how humanity can be promoted without prejudice to scientific and medical aims"

W.M.S. Russell and R.L. Burch (1959)

Abstract

The use of animals in laboratories is governed not only by legislation, but by ethical guidelines. Indeed, these ethical guidelines have been incorporated into legislation and underpin best practice to achieve good welfare and quality of scientific process. This chapter discusses the three guiding principles of humane experimental technique, the 3Rs: Replacement, Reduction and Refinement. There exists some evidence of a link between good welfare and good quality of data output in other laboratory-housed species, but not the dog. The principles of good science and ensuring high quality of data output are discussed, as are Refinements which have been shown to positively affect data output.

2.1 Ethical considerations on the use of dogs in laboratories

While there are many reasons to promote good welfare in the laboratory animal, the first of these should be ethical. The uses of animals in procedures and studies which have the potential to cause pain, suffering, stress or lasting harm are predominantly for human benefit. It is morally right to minimise any harm to the animals involved.

As well as ensuring that poor welfare does not compromise scientific objectives, and that good welfare promotes good science wherever possible, scientists working in the interests of the public, especially those funded by the public, have an obligation to conduct humane research as demanded by those who will ultimately benefit from the research. Surveys of public opinion consistently show that support for research using animals is greatest when the human benefit is greatest and that animal suffering is reduced as far as possible (Ipsos MORI, 2012). This is reflected in legislation, such as A(SP)A (1986), which requires that "cost-benefit" analyses be performed to justify the use of animals. Increased public awareness of the welfare of laboratory animals has also led to lobbying by scientists, and others, of the European Union and the emergence of a new directive mandating higher standards across the EU (2010/63/EU).

A utilitarian approach is often adopted in the assessment of costs, or harms, and benefits in animal research (Prescott, 2010). The principle of utilitarianism is that the benefit of an action (its utility) is measured in the 'happiness' it provides to living beings (Bentham & Mill, 2004). The underlying question which drives ethical debate in the use of animals for scientific purposes is *are the harms caused to animals justified by the benefits to humans?* There are varying approaches taken to assessing the harms and benefits, caused by the varying value attached to human gain and animal suffering (Nuffield Council on Bioethics, 2005). The evaluation of harms and benefits has been incorporated into frameworks and legislation, from A(SP)A (1986) to the incoming Directive 2010/63/EU, which introduces animal welfare panels for this explicit purpose. In the UK, the Animal Procedures Committee's *Review of the cost-benefit assessment in the use of animals in research* (Animal Procedures Commitee, 2003) provides guidance in addition to A(SP)A. The purpose of bringing this analysis into legislation is to ensure there is always a reasoned, justified case for the use of animals in harmful scientific procedures and to base decision-making not on feeling, but on considerations of ethics and science, both animal welfare science, and the best models of biomedical science.

2.1.1 Ethical guidelines for using dogs in laboratories

The three 'R' principles (Replacement, Reduction and Refinement) of humane science, which have become central to legal and ethical frameworks governing animal use, were first described by Russell and Burch (1959). Russell and Burch postulated that considerations of animal welfare was central to good science, and did not require compromise in scientific aims, a less contentious view now. Indeed, "humanity" as they called it, is integral to good science,

"*If we are to use a criterion for choosing experiments to perform, the criterion of humanity is the best we could possibly invent... The greatest scientific experiments have always been the most humane and most aesthetically attractive, conveying that sense of beauty which is the essence of science at its most successful*" (pg. 157).

These principles were central to informing the reduction of animal use and improvements in animals' welfare at a time in which there was much investment and many advances in the biomedical sciences (Richmond, 2010). However, the use of the 3Rs has evolved as changing scientific knowledge has developed, and therefore the definitions for the 3Rs has changed from Russell and Burch's original definitions. The UK's National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) has adopted updated definitions of the 3Rs which are currently the operational definitions (NC3Rs, 2010).

There are some key differences in these sets of definitions. Advances in research methods have allowed for the use of computer modelling and in vitro methods as suitable Replacement methods, and while Russell and Burch's definition prohibits the use of conscious living vertebrates, the NC3Rs definition recognises the place of non-protected animals and human volunteers in Replacement methods. The NC3Rs also gives greater breadth to the definition of Reduction, stating that not only should the numbers of animals be reduced to the minimum, but also that greater information can be gained from a given number of animals, thereby reducing future need for animal use. The greatest difference, however, is in the definition of Refinement, applied simply to the harms caused by procedures by Russell and Burch. This latest definition acknowledges that Refinement should be applied to many aspects of the animals' experiences, and encompasses not only reducing negative welfare but also increase positive welfare.

Table 2.1: Evolving definitions of the 3Rs

	Russell and Burch definition	**NC3Rs operational definition**	**Updated theoretical definition**
Replacement	Any scientific method employing non-sentient material which may in the history of animal experimentation replace methods which use conscious living vertebrates	Refers to methods which avoid or replace the use of animals defined as "protected" under the UK Animal (Scientific Procedures) Act 1986. These can be absolute replacements (e.g. computer modelling, in vitro methods (using human cells), human volunteers) or relative replacements (e.g. invertebrates, such as fruit flies and nematode worms)	
Reduction	Minimising the number of animals used to obtain information of a given amount and precision	Refers to methods which minimise animal use and enable researchers to obtain comparable levels of information from fewer animals or to obtain more information from the same number of animals, thereby reducing future use of animals (e.g. improved experimental design, modern imaging techniques, sharing data and resources)	
Refinement	Measures leading to a decrease in the incidence or severity of inhumane procedures applied to those animals which have to be used	Refers to improvements to husbandry and procedures which minimise pain, suffering, distress or lasting harm and/or improve animal welfare (e.g. environmental enrichment to improve the living conditions of research animals, anaesthesia and analgesia for pain relief, non-invasive techniques)	Any approach which avoids or minimises the actual or potential pain, distress and other adverse effects experienced at any time during the life of the animals involved, and which enhances their wellbeing

The definition of Refinement has continued to evolve from Russell and Burch's original concept to "reduce to an absolute minimum the amount of stress imposed on those animals that are still used"(Russell & Burch, 1959, p. 64). While Russell and Burch's original use of the term "Refinement" was applied to improvements in procedures and husbandry, and only considered decreasing negative welfare, Buchanan-Smith et al. (2005) proposed a new definition (see Table 2.1), taking into account all experiences throughout the life of an animal from birth to death in a more holistic approach. Other relevant aspects of an animal's life can include breeding and early rearing environment, weaning, sourcing and transport, housing, and the eventual endpoint for the animals. Where these have a negative impact on welfare, they are known as contingent harms, Refinement ensures that all aspects of the animals' lifetime experience are taken into account. Most importantly this recognises that welfare may not only be diminished by direct harms such as scientific procedures. Furthermore, it is critical that all animals are included in this definition; it is not only the animals destined for use in procedures which may suffer negative welfare, and Refinement should also be applied to founder animals in breeding colonies, thus encompassing all aspects of animal use within scientific research (Buchanan-Smith et al., 2005) . Many of the aspects of the laboratory environment with the potential to cause distress to dogs are not regulated procedures and to omit these from the sphere of Refinement would be negligent. Poor welfare in the breeding facility would likely predispose the dogs to poor welfare in later life and influence coping styles (see Chapter 1). The need to enhance good welfare is once again highlighted, rather than simply minimising harms.

The 3R principles underpin current legislation: e.g. A(SP)A (Home Office, 1986), legislation used by the Home Office to regulate research using animals in Great Britain and European Directive 2010/63/EU (European Union, 2010), on the protection of animals used for scientific purposes. Indeed it is incumbent on researchers to demonstrate that there is no alternative to animal use, that the number of animals will be kept to a minimum and that suffering will be minimised. In practice, the 3Rs are critical guiding principles to ensure we meet our ethical and scientific obligations in the use of animals for scientific purposes. This is especially true given the implications for the validity of the science as housing, husbandry and use in scientific research affects physiology, immunology and behaviour. Moreover, Robinson (2005) argued that applying the three 'R' principles is vital in scientific research to satisfy our society's moral obligations to reduce suffering.

2.2 The scientific importance of dog welfare in laboratories

Quality of science is inherently valuable, and obtaining the best possible quality from any scientific endeavour should always be a goal of those conducting research. The term "quality of science" may be interpreted in several ways and the following sections describe how it is interpreted in relation to the scientific use of animals. In this instance, "quality of scientific process" refers to the manner in which work is designed, conducted, analysed and presented, while "quality of scientific output" refers to the data output obtained. Quality of science can be thought of as the product of both of these factors. We can consider that there are two aspects to good science: quality of scientific process and quality of scientific output.

Furthermore, Poole (1997, p. 117) argues "good science" meets three central criteria, namely:

1. "There is an important question for which an answer is sought" (validity);
2. "The experiment should yield unambiguous results" (robustness);
3. "Variables which are not under investigation are strictly controlled" (robustness/reliability).

It should be easy to see how each of these factors can be influenced by a desire to ensure high welfare. It is explicit in the legislation governing animal use (Home Office, 1986) that to justify the use of animals, a study must have undergone a cost-benefit analysis and that the potential benefit from the information obtained (to science or society) must sufficiently justify the associated animal suffering. The study design must also be capable of obtaining the desired results. This means that the result should not be biased by poor study design or interpretation of the data. Control of extraneous variables is intrinsically linked to producing unambiguous results. Extraneous variables which are not anticipated to have an influence on results, biasing data output, resulting in incorrect interpretation of data. Tables I1-I3 in Appendix I lay out some of the ways in which we can understand the concept of 'quality of science' and the factors which influence 'quality of data'.

Poole also stated that:

"*Most scientists working with animals will make the assumptions that they will have normal blood pressure, heart rates, levels of stress hormones, immunological*

competence, digestion, appetite and behaviour. To avoid confounding variables, experimental animals should have both normal physiology and behaviour." (pg. 177)

Normal physiology, in this case meaning normal biological functioning rather than normally-distributed data, may not be present or properly understood (Poole, 1997). Assuming this in a situation where it is not true leads to poor quality of data, and indeed poor quality of science by designing a poor study with little chance of providing unambiguous data. The following sections discuss the existing evidence for a link between welfare and quality of science.

2.2.1 Linking welfare and quality of science

Animals are used as models for humans in research. Although it is never possible to say that a model responds to a treatment in exactly the same way as a human would, it is important to choose models which, as far as possible, predict that the response of humans. The first step in choosing a model is ensuring it is relevant to the target species, and the second is ensuring that the experiment is capable of detecting responses in the model (is sensitive to treatment effects, Festing, 2010). The validity of a model depends on how closely the model resembles humans for the specific characteristic being tested (Festing & Altman, 2002). High fidelity models are those which which resemble the target (in this case humans). Nonhuman primates are a prime example of a high fidelity animal model owing to their close relatedness to humans (Festing, 2010). However, other organisms or models such as cell cultures may model a specific human system closely despite their lack of fidelity for the human as a whole, they are high fidelity for that particular system.

Gad (2006), in a manual on animal models in toxicology agrees that the effects of stress and other biological responses are amongst the "least unaccounted for variables" in laboratory animal science (pg 852). While physiological responses have been identified as influencing responses in toxicology testing, they are rarely factored into experimental design or analysis. For example, Tasker (2012) found that restraint had considerable impact on many key toxicology measures in macaques. Everds et al. (2013) provides detailed information on the systems and measurements which are likely to be influenced by stressors, reproduced in Table 2.2. Everds et al. (2013) provide evidence of stressors affecting many of the organ systems key to safety assessment; unaccounted for, these effects of stress could bias the interpretation of results considerably.

Scientific progress is driven by developing and testing novel hypotheses and appropriate and robustly designed experiments are fundamental to this process

TABLE 2.2: In-life and pathology parameters routinely evaluated in toxicity studies: typical responses to stress, adapted from Everds et al. (2013)

Stage of life	**Affected system**	**Parameter**	**Potential stress effect**
In-life	Nutrition	Body weight or body weight gain	Decreased
		Food consumption	Decreased
	Circulating blood cells	Eosinophil counts	Decreased
		Lymphocyte counts	Increased or decreased
		Neutrophil counts	Increased
	Immune	Macrophage phagocytosis	Decreased
Organ weight	Endocrine	Adrenal gland	Increased
	Immune	Thymus	Decreased
		Spleen	Decreased
	Reproductive	Testis	Unchanged (rats) or decreased (mice)
		Epididymis	Decreased
		Seminal vesicles	Decreased
		Prostate	Decreased
		Ovaries	Decreased
		Uterus	Decreased
Organ histology	Digestive	Gastric ulceration	Increased
	Lymphoid	Thymic cellularity	Decreased
	Reproductive	Testis	Unchanged (rat) or degeneration (mice)
		Epididymis, prostate, seminal vesicles	Possible atrophy
		Ovary, uterus	Inactive
		Vagina	Atrophy, mucification
	Endocrine	Adrenal cortex	Hypertrophy/ hyperplasia

(Kilkenny, Browne, Cuthill, Emerson & Altman, 2010). Ensuring that studies are well-designed is not only important for ethical reasons, but also to ensure the best use of time, money and to further our scientific knowledge (Poole, 1997; Festing & Altman, 2002; Festing, 2010; Kilkenny et al., 2010).

Although it may seem obvious, excessively large studies waste resources (and most importantly animals), while those which lack power or have an element of bias may give the wrong answer, so adequate time should be dedicated to developing a suitable research strategy *a priori*, which may involve several individual experiments in order to ensure animals and resources are not wasted (Festing, 2010). While our ethical obligation to Reduce the number of animals used is often cited as a reason to ensure good experimental design, it must also be unacceptable to design studies which waste money, time and researchers.

Gad (2006) provides a list of the potential causes of animal studies not predicting the results of human trials (Table 2.3), for reasons relating to experimental design and welfare. Although Gad does not state explicitly that any of these reasons relate to welfare in the animal, there are clearly a number of reasons that animals can fail as experimental models for humans, excluding welfare. This only serves to highlight the importance of designing animal studies to achieve the best possible results, given that there are so many potential factors which can limit the ability of a study to detect the desired effect.

Building on Poole's principles of good science, Festing (2010) describes five fundamental characteristics of well-designed studies:

1. It should be unbiased with all subjects having the same environment unless the environment is the subject of the study. This can be achieved by randomisation of factors throughout the study, or use of factorial designs to determine the influence of environmental factors (robustness);

2. All experiments should have adequate power so that if there is an effect, there will be a high chance of detecting it. This can be achieved by controlling variation. Animals should be of the same sex, age, weight, health status and housed in the same environment as far as possible. Pathogens and disease increase variability and interfere with results. Stressed animals are also more variable physically and behaviourally. Randomisation should be used where it is not possible to control all factors. It may also be useful to take individual measures before beginning a study so that final measurements can be corrected for individual variation. Once variation has been controlled as far as possible, sample size can be determined with a power analysis (confounding factors);

TABLE 2.3: Adapted from Gad (2006) Reasons data obtained in animal studies may not always match human experiences

Issue	**Reason**
The animal species selected differ in response from humans. The same measurement or experiments in a different species might have been more precise	Validity, fidelity
Differences in absorption, distribution or metabolism might be present	Biological function or product of stress (see Everds et al., 2013)
The anatomy involved in the model might differ from that in people	Validity, fidelity
Different animal strains of the same species might generate different results	
The pathological nature of any lesions produced might differ at either a macroscopic or microscopic level	Robustness
There could be critical differences between the species at subcellular, cellular, receptor or physiological levels that lead to different responses. This is particularly true in terms of our current use of clinical chemistry findings to identify "target organs" in animals when those enzymes might not have the same relationship to pathogenesis in animals as humans	Fidelity
Experimental conditions in the animal model might yield qualitatively different data over the course of several experiments, and it might be unclear which set is	Potential interaction with welfare
The "dose" required to produce the observed results in animals is never achieved in humans	Robustness, sensitivity
The target dose in humans cannot be achieved in test animals	Poor model selection
The human population we are concerned about might differ from the population in general, and in so doing might have special characteristics that were not adequately represented in out animal model population	Poor experimental design

3. If it is important to know the effect of strain, sex, diet or other factors on the outcome, a factorial design should be used. This can result in greater information from the investigation of several variables and their interactions without the need for greater numbers of animals (robustness, validity);

4. Experiments should be simple so that chances of making a mistake are minimised. This means studies should be well planned in advance, with no additional components added at a later point as randomisation will no longer be possible (robustness);

5. The experiment should be amenable to suitable statistical analysis. The most important criterion in this case is independent replication of results. There should be a clear understanding of how the results will be analysed before beginning the study, with researchers consulting a statistician where necessary.

Clearly, if the experiment is not designed with sufficient power, a treatment effect may not be detected, resulting in a false negative (Kilkenny et al., 2010). The scientific method assumes the lack of confounding factors or uncontrolled variables (Poole, 1997) and so reducing variation is an important component in increasing power. This can be done by controlling genetic variation by using inbred strains. Although it is sometimes argued that inbred strains reduce external validity (e.g., Würbel, 2002), Festing (2010) states that it is false logic to use outbred strains because nothing is known about the genetics of the subject and this results in an increase in phenotypic variability, reduced power and reduced repeatability. Factorial or crossover designs[1] are particularly powerful, utilising within-subject and between-subject factors (Shaw, Festing, Peers & Furlong, 2002; Festing, 2010) and result in the need for fewer animals. These designs respectively control for genetic and environmental variation and illustrate the effects of genetic and environmental factors and their interactions.

Increased or decreased variance can be caused by infection, genetics, environment, age, sex, weight, welfare state and other unknown factors (e.g, Poole, 1997; Festing & Altman, 2002; Würbel, 2002) and reduces the power of an experiment to detect treatment effects. Techniques such as the use of inbred strains of mice reduce genetic variation, thereby increasing the probability of detecting a treatment effect on a specific genotype (Festing & Altman, 2002), and when a treatment needs to be investigated in several phenotypes, factorial design can substantially reduce the number of animals needed (Shaw et al., 2002). The use of genetically modified animals is uncommon in primates, and unknown in dogs, and so this is primarily a technique confined to early pre-clinical testing using mice or rats.

[1]Factorial design: consists of two or more factors, each consisting of discrete levels. Crossover designs allow the analysis of the effect of each factor, and the interactions between them, on the outcome

Festing and Altman (2002) also support the use of historical data, which when carefully used can reduce the need for larger samples sizes in current studies. Meta-analysis and use of contemporary controls may be necessary to ensure that historical data are valid for use in a current study but may prevent the need to repeat previously conducted research. Caution should be taken when comparing data from populations which differ in welfare states however (Hall & Everds, 2008). Data from animals housed under different conditions or experiencing differences in handling are unlikely to be comparable, unless such variables are factored into analysis (Richter, Garner, Auer, Kunert & Wurbel, 2010).

Festing (2010) states that the randomised controlled, double-blinded[2] clinical trial is the gold standard for nearly all experiments, so where possible and appropriate, these factors should be included in experimental design. One of the most important factors in experimental design therefore is to ensure that the data produced is affected only by the variables under investigation, or where other variables may influence output, that they are accounted for in experimental design and analysis. However many of these factors may appear to be unattended in contemporary research (Kilkenny et al., 2009), a factor which the ARRIVE guidelines (Kilkenny et al., 2010) seek to address.

2.2.2 Quality of data output

While there is clearly a link between the ability of an animal to cope with its environment and its physical and emotional health, it is all the more important to understand this link where the animal is a model for human subjects. Although many therapeutic drugs target specific areas of ill health, the desired animal model in toxicology is a healthy animal, rather than one with unknown, stress-induced physiological health issues. Without the ability to understand the specific variation introduced by poor welfare, it is not possible to have an adequately-designed experimental protocol, nor obtain valid results. Quantifying the effects of welfare on quality of data output is one of the overarching aims of this project, allowing the proper design of data collection to achieve the aims of studies. The following section describes how issues in quality of data output can be identified and improved.

Once again, we must return to Poole's 'happy animals': to ensure good science in research using animals, the animal subjects should have biologically normal physiology and behaviour; animals whose 'wellbeing' (or welfare) is compromised are often physiologically abnormal and the results of experiments using them may not be reliable (Poole, 1997). The emotional, subjective experience of animals (Chapter 1) is

[2]Double-blinded: neither those conducting nor analysing studies are aware of the allocation of animals to treatment conditions

central to their welfare and as such should be considered central to their use as experimental models.

As we have already discussed the criteria identified as being important to quality in scientific research using animals, it is now appropriate to review current research to assess whether these criteria are being met, and to determine where improvements can be made. Several reviews (Festing & Altman, 2002; Kilkenny et al., 2009; Festing, 2010) have stated that a review of recent research using animals illustrate that many of the principles of good science outlined in the preceeding section are not adhered to, which can result in the publishing of research with poor validity. Festing and Altman (2002) state that there are papers published in which the conclusions reached are not supported by the data.

Kilkenny et al. (2009) assessed the quality of current research using animals by analysing experimental design, statistical analysis and reporting of results in journal papers in a survey commissioned by the NC3Rs. The survey assessed 271 papers published between 1999 and 2005 reporting publicly-funded original research on live rats, mice and nonhuman primates, as these constituted the greatest part of the literature on research on live animals. Less than half the papers reported the age (43%) or weight (46%) of the animals used, while 24% reported neither. A small percentage (4%) did not report the numbers of animals used and no paper reported how the number of animals needed was decided. The characteristics of the animals used influences the results obtained and is required to replicate experiments. The number of subjects is important for statistical analysis and significance, and the decisions which lead to the number of animals used should be scrutinised to ensure that the 3Rs have been adhered to.

Further analysis found that 35% of papers reported different numbers of animals in the methods and results sections without clear explanation of the difference. In all, only 59% of papers reported a clear hypothesis, at least three of: animal sex, strain, weight and age, and the number of animals used. This reflects a lack of understanding of the importance of reporting key experimental details to ensure quality of science and transparency on the part of researchers. Transparency, as well as the ability to understand and reproduce research, is all central to good scientific method.

The authors also assessed the quality of experimental design. Random allocation of animals is a process used to ensure that as far as possible, differences in outcome measures cannot be attributed to random variation and is concurrent with Festing's "gold standard". Only 12% of papers reported the use of random allocation. Blinding is a method of minimising bias by ensuring that the experimenter does not know to

which condition a subject is allocated. This is important when subjective measures are used. Of the papers using qualitative measures, only 14% used blinding.

Factorial design allows combinations of two or more factors to be evaluated in one experiment, maximising the amount of information gained from a sample of subjects. This can reduce the number of animals needed (and also increases the validity of results). Of the papers in this report assessed to be suitable for factorial design, only 62% utilised it. This survey found problems with analysis and/or the reporting of analysis in 60% of papers, while only 70% actually described the statistical method used and provided a measure of error or variability.

In addition to the factors which can prevent an animal model from accurately predicting human responses in trials (Table 2.3), without proper experimental design (Section 2.2.1) or correct reporting of animal characteristics and analysis methods, it is difficult to determine if experimental results are valid. In the scientific use of animals for the pursuit of new medicines, when we use animals in studies with the capacity to cause pain, suffering, distress of lasting harm (Home Office, 1986), and when the end-users of the test items under investigation are the public and health care providers, it is critical that the best possible quality of scientific investigation is adhered to. Details of housing and husbandry are relevant to the interpretation of any study findings.

In the interests of maintaining this quality of scientific inquiry, the reporting of the results of a study must be done correctly. To be considered valid, results must be replicable, which necessitates that sufficient detail must be provided. Kilkenny et al. (2009) state that there is a responsibility on the part of journals and the peer community to ensure that published research is of a suitably high standard and is reported correctly and transparently. As a result of the Kilkenny et al. (2009) investigation, the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) Guidelines, as published by Kilkenny et al. (2010). More recently, a guide to the implementation of key principles of Good Statistical Practice (GSP) was published by (Peers, South, Ceuppens, Bright & Pilling, 2014). This included standards in statistical practices, identification of responsibility for adhering to GSP, improvements to report writing and ensuring decisions are made data driven.

2.2.3 The link between welfare and quality of cardiovascular data

One of the predominant uses of dogs other than toxicity[3] in safety assessment is in the assessment of the effects of a compound on the cardiovascular (CV) and respiratory

[3]Toxicity: degree of damage caused to organ systems

systems (Koerner & Siegl, 2013). The CV system of the dog more closely resembles that of humans than does that of rodents', and dogs often replace the use of nonhuman primates in this type of research because of the aforementioned availability of historical data and apparent similarity of response (Chapter 1). There are several aspects of CV function which are of particular interest in safety testing. These are blood pressure[4] (increases or decreases), heart rate[5] (increases or decreases), heart rate variability[6] (decreases) and QT interval[7](increases). As is set out in the following sections, changes in these parameters, particularly in individuals with pre-existing cardiac conditions, are considered to be highly undesirable.

Mitchell et al. (2010) state that the gold standard for cardiovascular safety evaluation of new drugs is the use of radiotelemetry, primarily for safety pharmacology studies in large animals such as non-human primates and dogs. Handling and restraint can have profound effects on the data obtained and telemetry negates the need to handle and sometimes restrain animals during data collection. When high quality data can be obtained from telemetered devices, this can contribute to Reduction, while the ability to observe animals in the home pen reduces the disturbance associated with removal from the pen and handling for electrocardiogram (ECG) readings, which in turn contributes to Refinement and to the quality of data obtained (Hawkins et al., 2004). Longer term toxicology studies need repeated dosing to evaluate the safety of small and large molecules and therefore need accurate and reliable non-invasive techniques for measurement of CV function. Methods should be not be invasive as something which directly interferes with the cardiovascular system may interfere with interpretation of pathology end points (e.g. organ weights, blood chemisty). Refinements to the acquisition of cardiovascular data include the use of jacketed telemetry (e.g. Chui et al., 2009), which requires acclimatisation to the equipment but does not require surgery and is not in itself considered a regulated procedure. Other methods of obtaining ECG such as readings taken from gently restrained conscious dogs may be appropriate in long term studies, with an initial investment in dog training required to mitigate the effects of restraint stress on data obtained.

Measures such as heart rate, blood pressure and QT interval are used in the studies comprising this project, both to integrate cardiac measures with behavioural measures of welfare and to investigate the effects of changing welfare on these measures in order to investigate quality of science.

[4]Blood pressure: usually given as the minimum and maximum pressures exerted on the arteries during the cardiac cycle, measured in millimetres of mercury (mmHg)

[5]Heart rate: the number of heart beats in one minute, measured in beats per minute (bpm)

[6]Heart rate variability: the relationship between high and low frequency heart beats, reflective of sympathetic nervous system activity

[7]QT interval: time in milliseconds between Q and T components of the ECG; a key safety measure

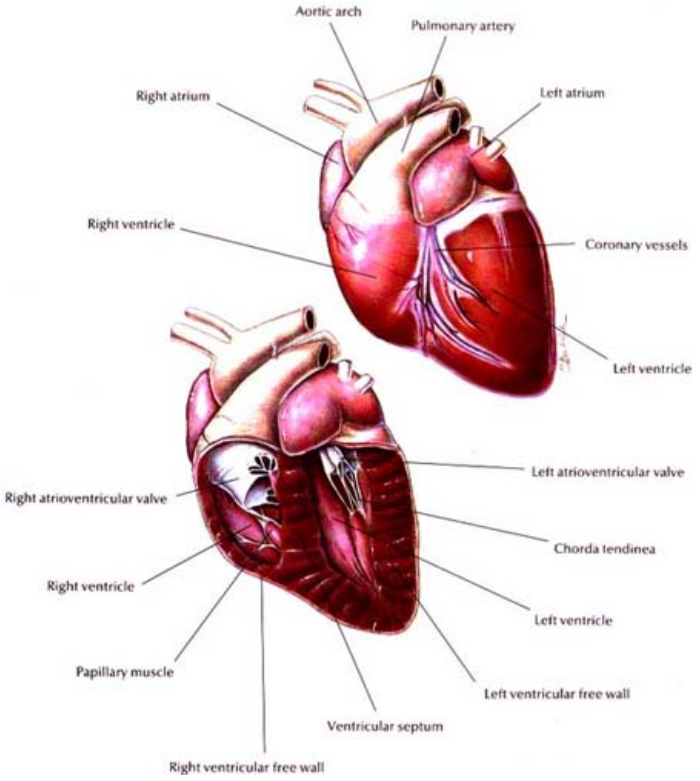


FIGURE 2.1: Canine heart - normal anterior view (Hills Pet, accessed 2014)

2.2.3.1 The canine heart

The heart is a vital organ, located in the chest cavity surrounded by a protective sac (the pericardium). The heart is a muscular organ which pumps blood around the body. Deoxygenated blood enters via the superior and inferior venae cavae into the right atrium and is pumped to the lungs via the right ventricle and the pulmonary artery. Oxygenated blood re-enters the heart via the left atrium and pulmonary veins, and is pumped to the body via the left ventricle and the aortic artery.

The cardiac cycle comprises five stages (Gross, 2009): 1: atrial systole (atria contract); 2: isovolumetric contraction (ventricles begin to contract); 3: ventricular ejection (pressure in ventricles rises and blood is ejected into aorta and pulmonary artery); 4: isovolumetric relaxation (ventricles relax and valves close) and 5: ventricular filling (ventricles passively fill and atria expand). The two heart beat tones heard during the cardiac cycle are created by the closing of the valves. The first is caused by the closure of mitral and tricuspid valve while the second is caused by the closure of the aortic and pulmonary valves. Electric activity (contractions) are controlled by the sinoatrial (SA) and atrioventricular (AV) nodes (Gross, 2009). Electrical signals pass from these nodes into the heart causing contractions in the cardiac muscle. The full cardiac cycle can be seen in Figure 2.3.

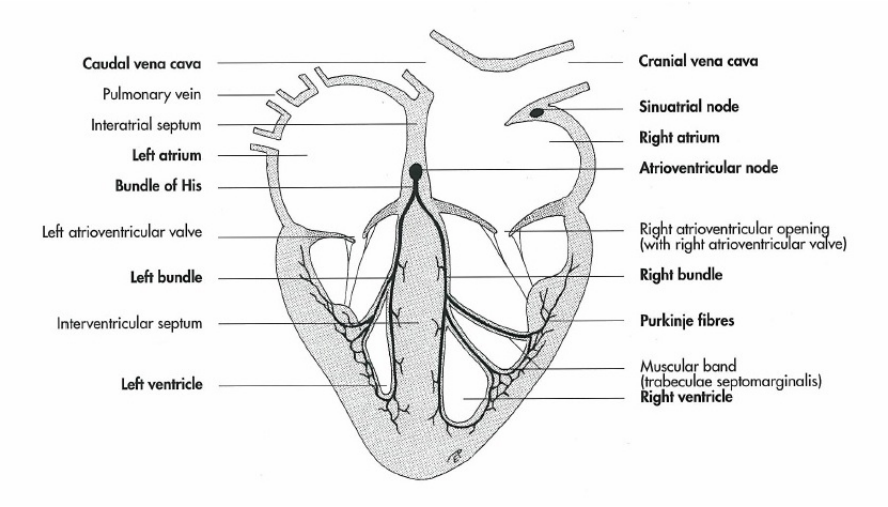


FIGURE 2.2: Conducting system of the right and left ventricle, schematic (König et al., 2004)

The cardiac cycle as depicted in an electrocardiogram (ECG) is shown in Figure 2.3. The heart beat is defined as the the number of R-R intervals in a 60 second period. QT interval describes the time period between the Q and T waves on an ECG, and reflects electrical conductivity; while T-wave morphology refers to the particular shape of the T-wave. The interval represents the polarisation and depolarisation of the nodes which control the contractions of the heart muscle. Particular attention is paid to the length of the QT interval in safety assessment as lengthening of the QT interval can indicate problems leading to *torsades de pointes* (seen as a rotation around the horizontal axis) or sudden death in humans. This makes QT interval an important measurement in safety assessment (Valentin, 2010) QT interval increases or decreases with heart rate, therefore in order to investigate the QT interval in isolation, correction factors are applied (see Chapter 5). The most commonly applied correction is Bazett correction (Bazett, 1920):

$$QT_B = \frac{QT}{\sqrt{RR}} \tag{2.1}$$

where QT_B is the corrected interval (ms), QT is the QT interval and RR is the R-R interval (ms), however (Tattersall, Dymond, Hammond & Valentin, 2006) found that Van der Water's correction (below) was a statistically superior correction, and that correction factors should be developed by each facility based on the strain of dog and technique used. This is the formula applied in Chapter 5 of this thesis.

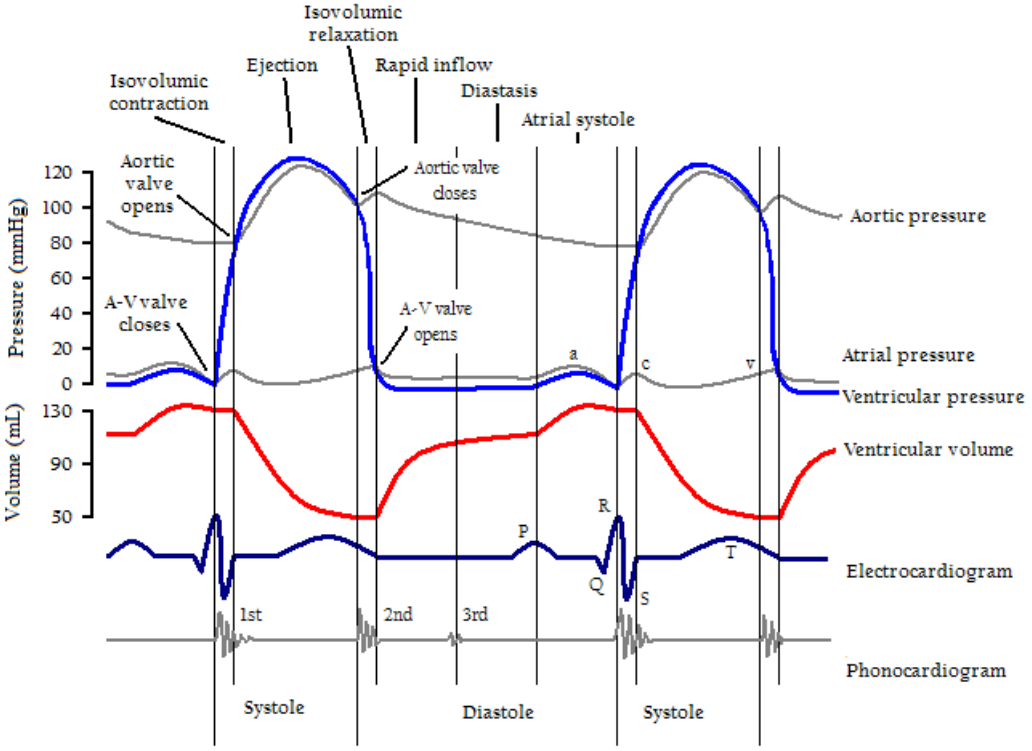


FIGURE 2.3: A Wiggers diagram, showing the cardiac cycle events occurring in the left ventricle (Chang, 2011). The relationship between blood pressure and other cardiac events can be seen. In the electrocardiogram: wave "P" corresponds to the onset of atrial depolarization, waves "QRS" correspond to the onset of ventricular depolarization, and wave "T" corresponds to ventricular repolarization. The phonocardiaogram shows the two heart sounds in relation to the ECG.

$$QT_{VdW} = QT - 0.087(RR - 1000) \tag{2.2}$$

Accurate recording of and correct interpretation of ECGs is important to assess the effects of drugs on cardiac repolarisation and risk of arrhythmia in humans. ECG quality can be influenced by factors such as motions artefacts (movement of the subject) or interference (in the case of telemetered data). QT and changes in T-wave morphology are considered critical parameters to monitor in toxicology or pharmacology studies (Hanton & Rabemampianina, 2006).

It is important to understand the factors which might influence ECG parameters such as sex, body weight, genetics and heart rate. HR affects PQ and QT intervals and possibly the amplitude of the P-wave (Hamlin, 2008). HR may be altered by drug treatments, experimental conditions and in particular stress and excitement. In safety assessment, it is necessary to distinguish between the effects of the drug on QT and

indirect effects such as the orientation of the dog or restraint stress (Hanton & Rabemampianina, 2006).

2.2.3.2 The venous system

The pressure of blood in the heart and venous system (measured as millimetres of mercury, mmHg) varies throughout the cardiac cycle. Arterial pressure is influenced by blood vessel contractility (or degree of elasticity), and with reducing ability to adapt to volumes of blood, pressure increases. Increased blood pressure is associated with a number of negative health outcomes in humans, with lower blood pressure conferring health benefits (MacMahon et al., 1990). A medicine which has the side effect of increasing blood pressure would be undesirable, especially in an already-vulnerable population and as such, blood pressure is central to the battery of measures used in safety pharmacology (Koerner & Siegl, 2013).

As well as arterial pressure, pressure in the left ventricle (LVP) is now frequently measured in safety assessment studies, as although only heart rate, blood pressure and ECG are required by ICH S7A/B (Leishman et al., 2012), late stage discovery of alternations in LVP is undesirable. During diastole, LVP is considerably lower than arterial pressure, but increases to the same level during systole as the ventricle contracts, the aortic valve opens and blood is pumped to the body. The pressure in the ventricle is related to its contractility, as reduced ability to accommodate increasing volumes of blood leads to increasing pressure (Gross, 2009). Increased LVP and reduced contractility means that the heart requires more oxygen to function, which can exacerbate existing heart disease. An increase in arterial pressure also means that the heart has to contract more forcefully to maintain the volume of blood being pumped (Sjaastad, Hove & Sand, 2010). Although LVP was present in the data collected from telemetered dogs (see Chapter 5), only arterial systolic and diastolic pressures, and heart rate, were analysed, given the link to physical health and welfare identified (see the following section).

2.2.3.3 The effects of physiological or psychological stress on the cardiovascular system

While it seems obvious that physical exertion can have a profound effect the cardiovascular system, psychological stress can have a similar effect. The effects of psychological stress may go undetected however, without attention to behavioural indicators of stress. There is evidence that sudden increases in heart rate and blood pressure are more damaging than chronic increases over a period of time, as a result of

the heart having insufficient time to adapt to the change in function. In humans, an athlete who regularly undergoes cardiac exercise will be better able to cope with sudden increases in heart rate and more quickly able to return to baseline levels than someone who never exercises. Langer, Obrist and McCubbin (1979) found that in rats, the psychological stress associated with learning to avoid shock (using operant conditioning) produced a similar cardiovascular response as undergoing intermediate or heavy exercise.

In the dog, where laboratory dogs lead a relatively sedentary life (Hubrecht et al., 1992) and are subject to potentially distressing events which cause sudden increases in heart rate and blood pressure, there is a concern that subtle damage is being caused that is not otherwise accounted for in scientific design.

There are ways to measure the effects of stress on the cardiovascular system; cardiac troponin (particularly types I and T) are proteins which are used as sensitive indicators of damage to the muscle of the heart. Damage results in release of troponin into the blood stream, which can then be measured. In humans, this is considered the gold standard test for a myocardial infarction

Van Citters and Franklin (1969) conducted research using chronically telemetered sled dogs to investigate extreme fitness and cardiac stress. When resting, the dogs' heart rate was typically 40-60 bpm,and blood pressure was 100/70-150/100 mmHg. During pre-race excitement, heart rate increased to 120-150 bpm. During the exertion of racing, heart rate was around 250-300 bpm, maximum recorded blood pressure 'occassionally exceeded 300 mmHg' and diastolic run off (transfer of blood to peripheries) was rapid to account for this. However, mean heart rate was approximately the same during exertion as it was at rest, suggesting that sufficient adaptation has occurred in the heart to function at the higher heart rate without causing damage associated with high blood pressure. Although no absolute maximum value is available for the dog, these values appear to be the highest achievable by the most physically athletic dogs; maximum values in less fit dogs are likely to be considerably lower. Similarly, Vatner, Higgins, White, Patrick and Franklin (1971) subjected 'normal' dogs to severe exercise (running at 25 mph) and found that heart rate increased from 84-259 bpm, while blood pressure increased to around 140 mmHg from 89 mmHg (no diastolic value given). This suggests that athletic dogs are better adapted to sudden increases in cardiac activity associated with severe exercise. Van Critters and Franklin's dogs' heart rate recovered to 150 bpm with 1 minute and then rapidly to baseline, while Vatner et al. (1971)'s dogs took 45 minutes to recover. In humans, it is known that sudden increases in heart rate and blood pressure can be dangerous (Prisant, Carr & Hawkins, 1993) and there is no reason to believe this is

Table 2.4: Classification of blood pressure (BP in mmHg) in dogs based on risk for future target organ damage (TOD), adapted from Brown et al. (2007)

Risk Category	Systolic BP	Diastolic BP	Risk of future TOD
I	<150	<95	Minimal
II	150-159	95-99	Mild
III	160-179	100-119	Moderate
IV	>180	>120	Severe

different in the dog. It is also known in humans that a small increase in blood pressure can be detrimental to health with a 2 mmHg increase resulting in a 10% increase in stroke risk and 7% increase in mortailty (Lewington, Clarke, Qizilbash, Peto & Collins, 2002). It is well recognised that in humans, raised blood pressure has a negative impact on organ systems in the body, as well as contributing to mortality. It is important to understand these effects in a canine heart which is acting as a model for the human heart. As with humans, values have been established for dogs to reflect the impact of set ranges of raised blood pressure. The following table is adapted from Brown et al. (2007) and shows a range of systolic and diastolic blood pressure and the associated risk of target organ damage, specially vital organs such as heart, lungs and brain.

As in humans, there are 'maximum' values for arterial pressure which are rarely exceeded, the human value being 300 mmHg, (MacDougall, Tuxen, Sale & Moroz, 1985) above which the ventricle cannot contract, except in the case of highly fit athletes. Blood pressure above 180 mmHg is considered a 'hypertensive emergency' (Zampaglione, Pascale, Marchisio & Cavallo-Perin, 1996). Secondary hypertension, caused by an underlying disease has greater potential to cause target organ damage, as does acute hypertension. Unlike chronic hypertension, the body has not been able to develop adaptations to cope with the increase in blood pressure.

Hanton and Rabemampianina (2006) established reference values for ECG parameters of beagles, while investigating the effect of restraint method and breed strain on the values. Dogs from Marshall Farms, USA (n=1880) and Harlan, France (n=57) were studied. All were between 13-20 months old and were acclimatised to their surroundings for two months before the study began. Two methods of restraint were compared: suspended in a hammock or standing on a table, being gently restrained while ECG readings were obtained. The cardiac axis (rotation of the ECG reading around its axis) of the suspended dogs had a left shift while standing dogs had longer QT intervals. There were also significant differences in P-wave amplitude. The authors suggested that neither method seemed to give more reliable results however one method of restraint should be maintained within and between studies to prevent differences in parameters resulting from changed position. The major difference between strains was in mean and max PQ interval.

Variation in heart rate caused by factors unaccounted for (such as restraint) needs to be considered, as it was found that about half the variation in QT interval can be accounted for by variations in HR, suggesting that other factors were also influencing QT interval. The authors found differences in baseline measurements of PQ interval and P-wave amplitude, supporting the need for each strain of beagle to have its own 'normal' range of values. Similarly, Moscardo et al. (2009) examined cardiac disturbance in telemetered mature male beagles obtained from Marshall Europe. It was found that dogs familiarised with procedures were more relaxed and cooperative, expressed more normal behaviours, and this allowed examinations to be completed more quickly. Increases over baseline values in heart rate and systolic blood pressure (no exact values given) persisted for five minutes after the stressor of handling occurred, and were of longer duration in the last dog to be examined.

2.2.4 Harmonising welfare and quality of data output through Refinement

The concept of "harmonising" welfare in animals is one of providing all animals with the necessary tools to cope with the environment, in order for them to exhibit the same "harmonised" level of positive welfare (Buchanan-Smith, 2006). Individual differences may mean that animals have varying needs, so providing a variety of Refinements increases the ability of all animals to cope. The previous sections report a number of studies which have found that rather than increasing variation, increasing enrichments and other Refinements decreased the level of variation in the population. The reason for this is that individual differences in coping styles and abilities to cope vary, and providing the greatest possible variety of coping strategies increases the ability of the animal to manage stressors in its environment.

There are many aspects of the laboratory environment with the potential to decrease the welfare of laboratory-housed dogs. The following sections describe features of the environment which can be modified to have a positive impact on welfare.

While we have discussed how quality of science can be increased through improvements in experimental design, analysis and publication, there is also an obvious role for welfare. It is widely accepted that applying the 3Rs to experiments using animals is consonant with good scientific practice (Kilkenny et al., 2010). While laboratory animals do not lack essential needs like food or water, potential causes of distress include social problems including aggression resulting from overcrowding, social isolation, loud sudden noises and poor handling (Poole, 1997). The needs of the dog in

conspecific and human contact (Chapter 1) may not be met in the laboratory environment.

Festing (2010) have stated that stressed animals are more 'variable' than unstressed animals, and that disease and pathogens can interfere with experimental outcomes. However, positive changes in stress responses can change the outcomes of diseases, often beneficially. In a review, Van Praag, Kempermann and Gage (2000) cite the examples of slower neural degeneration, faster recovery from brain damage and improvements in HPA responses to stress, in experimental studies of invertebrate, rodent and human models of brain damage. Minimising stress during experiments can reduce variation (and therefore the number of animals required) although a thorough understanding of the animal and its biology are needed in addition to experiments which are well designed and statistically valid and appropriate (Poole, 1997).

Animals have evolved a range of coping mechanisms to natural stressors, including changes in behaviour, hormones or immune function. However, in captivity where the environment does not allow an appropriate coping response or overloads it, the animals ability to maintain homeostasis breaks down and leads to a state of distress (Hubrecht, 2010). Poole (1997) states that doubting the role of behaviour in understanding physical wellbeing is a result of misunderstanding that brain and body are linked. The brain, behaviour, hormones and the immune system are linked and interdependent. Several studies have found that stress has negative effects on physical health and the immune system. Examples of these findings are summarised in Table 2.5.

When considering the effects of stress, environmental enrichment has a clear role in influencing welfare and therefore experimental outcomes. Enriched environments provide more opportunities for animals to make choices, increasing their ability to maintain homeostasis or to control social interactions (Hubrecht, 2010). A number of authors reported identified concerns from scientists conduction research with animals that increasing the variability in the environment through increasing enrichment results in a loss of standardisation (Wurbel, 2001, 2002; Wolfer et al., 2004; Benefiel, Dong & Greenough, 2005; Hubrecht & Kirkwood, 2010). Other concerns include the cost of implementing enrichments, bias of experiments and risk to the animal (Hubrecht, 2010). The effects of changes in the environment seem to be most pronounced, or at least most readily detected in the development of the brain in mammalian species. Table 2.6 lists some examples of the influence of environmental enrichment on the brain.

TABLE 2.5: Examples of the negative effects of stress on health

Study	Species	Findings
Marsh et al. (1963)	Macaques	Increased resistance to polio virus in subjects exposed to avoid shock
Reite (1987)	Macaques	A two week separation from littermates at six months old resulted in lower T-cell proliferation in response to an antigen six years later.
Keller et al. (1981)	Rats	Increased rate of tumour occurrence in response to inescapable electric shock
Damon et al. (1986)	Rats	When allowed 21 days to acclimatise to metabolism cages, toxic dose of uranium was 220-650mg/kg, compared to 8mg/kg for rats not allowed to acclimatise.
Clough (1988)	Rats and mice	Bright lights result in retinal degeneration, raised prenatal mortality and decreased growth rates.
Brayton (1974)	Mice	Overcrowded mice had reduced resistance to a parasite
Blecha et al. (1982)	Mice	Restraint in a wire cone for two hours resulted in increased corticosteroid levels and immune response suppression.
Beden and Brain (1982)	Mice	Subordinate or defeated mice have a reduced immunological response to an antigen.

2.2.4.1 Predictability

The predictability of an event affects an animal's response to it (Weinberg & Levine, 1980, cited in Bassett and Buchanan-Smith (2007)). Predictability has been manipulated in two ways: *temporal predictability* - in which the stimulus is delivered on a fixed-time or variable-time schedule; and *signalled predictability* - in which the stimulus is preceded by a signal (Bassett & Buchanan-Smith, 2007). Animals have been shown to prefer shocks with high signalled predictability (Badia, Harsh & Abbott, 1979), with rats choosing longer and more intense predictable shocks over unpredictable shocks (Badia, Culbertson & Harsh, 1973). In a review, Bassett and Buchanan-Smith (2007) found conflicting evidence of the effects of predictability, with some authors concluding that predictability decreases stress and others concluding that it increases stress. Abbott, Schoen and Badia (1984) concluded that while an unpredictable stimulus may increase stress in the short term, it may be less stressful in the long term. This is because in the short term a predictable stimulus allows the animal to have 'safe periods' however in the long term this leads to heightened arousal while an animal will adapt to deal with the threat of unpredictable stimuli. This

Table 2.6: Examples of the positive effects of environmental enrichment on the brain

Study	Species	Findings
Volkmar and Greenough (1972)	Rats	Increasing levels of neurogenesis demonstrated through increased dendritic branching across isolated-, socially- and enriched-housed rats.
Turner and Greenough (1985)	Rats	Enriched rats had greater numbers of neurons and synapses, brain vascularisation and glial cells than isolation- and socially-housed rats.
Würbel (2001)	Rats	Increased enrichment results in structural changes, including increased numbers of neurons, synapses and dendritic branches.
Rampon et al. (2000)	Mice	Nonspatial memory deficits in genetically altered mice overcome by increased enrichment.
Healy and Tovee (1999)	Review of the literature	Brain size increases of around 5% in mammalian species as a result of increased enrichment.
Benefiel et al. (2005)	Review of the literature	Improved recovery from brain damage and slowing of neural degeneration as a result of increased enrichment.

suggests that for animals in laboratories, it may be beneficial to signal aversive events, at least when the events take place in the short-term. Research by Seery, Holman and Silver (2010) suggests that the relationship between exposure to stressors and ability to cope is influenced by both prior experience and level of adversity. Bassett and Buchanan-Smith (2007) make a number of suggestions for animals in a laboratory environment: an assessment of unintentional temporal signals of positive and negative events, negative events should be made temporally predictable; a reliable signal should be introduced to indicate the onset of a negative event; training using positive reinforcement may increase control and predictability for animals; and where possible, avoid delays following a signal to prevent a reduction in signal predictability.

When assessing husbandry and environmental variables which may be affecting welfare indicators and physiological measures, the ability of animals to predict and control events should be taken into account. Improving predictability and control may be an effective method to improve both welfare and quality of science (see Chapter 7 for an example).

2.2.4.2 Resilience

Increasing resilience to stress is an important aspect of improving welfare. Early experience has been shown to be important in an animal's later resilience. It is widely accepted that early (and positive) exposure to stressful situations in humans can increase resilience in later life through 'inoculation'. While many dogs will be obtained directly from breeders, the role of early experience in increasing resilience should not be discounted. Increased social conspecific enrichment was found to increase resilience in mice (as measured through the presence of anxiety-indicating behaviours in a stressful situation) in D'Andrea et al. (2010)'s study.

Pryce et al. (2005) reviewed the literature on early short- and long-term separations in rodent and monkey species and found increased acute cortisol responses (as well as behavioural responses) to various stressors in the long-term associated with increasing separations. However it should be noted that in most of the studies reviewed, the infants were isolated rather than being removed for brief periods of handling or socialisation.

Lyons, Parker, Katz and Schatzberg (2009) showed a similar effect in infant squirrel monkeys (*Saimiri sciureus*) separated from the natal group when later tested in a novel environment. These infant monkeys were separated from their natal group for one hour once a week for ten weeks. The authors state that this mimics natural conditions, however this is debatable. The infants did not exhibit high levels of distress during these separation periods, however exposure to this low level of stress seems to have had a positive effect on later ability to cope with a stressful situation, as indicated through their decreased behavioural stress response in a novel environment. Dettling, Feldon and Pryce (2002) also found that in marmosets (*Callithrix jacchus*), daily isolations on days 2-28 of life resulted in lower basal cortisol levels than controls when tested in a novel environment in weeks 18-20. However, infants in the experimental condition also exhibited fewer social calls during, and less contact with a family member after, the novel environment test. This suggests that the infants may have been coping by withdrawing from social contact, which is not otherwise conducive to good welfare. A decreased cortisol response may also be an indication of a dampened response caused by chronic stress, rather than lack of distress (Chapter 1).

It may be possible to increase resilience in dogs by gradual exposure to a stressor, or desensitisation, guidance for laboratory-housed animals is discussed by Laule (2010). Past experience influences perceived ability to cope with a stressful situation, as the animal has learned a set of problem solving strategies for the situation or if unable to cope may have learned helplessness. Social support can increase resilience by acting as

a stable base and also as an environmental buffer by providing a means of coping with stressors and reducing stress. For this reason, social housing is recognised as especially important in nonhuman primates and dogs (European Union, 2010). Increasing resilience can be observed in reduced impact of stressors, reduced recovery time following a stressor, growing competence at coping and an increase in capacity following each stressful exposure. An experimental evaluation of desensitisation protocols is presented in Chapter 7.

Blackwell et al. (2010)'s study outlines the relationship between coping with learning and physiological health. Their investigations into the stress responses of military working dogs suggested that those dogs most able to cope with exposure to an audio-visual stressor and learn a new task had higher levels of cortisol as well as more rapid rates of learning a new task. This was in line with the authors' hypothesis that stress may positively affect learning by narrowing focus and processing a smaller number of stimuli. Dogs which exhibited the lowest cortisol tended to be apathetic and did not interact with carers. Increasing the ability to cope with stress may result in dogs with more active coping strategies; this supports what we know about the stress response, and that by itself, cortisol responses to stimuli are insufficient to tell us about the meaning of an event to an animal.

2.2.4.3 Early experience

The importance of early experience for the laboratory-housed dog cannot be underestimated. The early months of a puppy's development are a crucial window in which socialisation with conspecifics and humans can have a life-long positive impact, or conversely, their absence can have a negative impact.

Puppies have limited organisational, sensory and motor abilities at birth (Fox & Stelzner, 1966). Their eyes and ears are closed and they do not respond to auditory or visual stimuli. However, they are born with developed tactile and thermal senses, responding to touch, pain, heat and cold (Jones, 2007). They also react to olfactory stimuli although this may be through taste rather than true olfaction. It is necessary to understand when and how well puppies can see, hear and feel to effectively implement any primary socialisation (Jones, 2007).

Jones (2007) makes the following recommendations based on the acceptance of critical periods of development. As experiences during critical periods may well have life-long impact on the dogs, it is important to monitor these experiences. Modification of whelping beds to minimise stress to newborns and increase tactile and thermal comfort may be beneficial. Gentle introduction of light and sound and these abilities emerge

may serve to desensitise pups to these. Desensitisation to unfamiliar stimuli during the period in which the startle response emerges may also be beneficial to the pups' future responses to unfamiliar stimuli.

One extensive study of the effect of early experience on later life was conducted by Wilsson and Sundgren (1998). A cohort of German Shepherd puppies (n=867) were followed from birth to six years. Puppy weight, litter size, temperature and bedding type were all found to influence later behaviour. The effect of puppy weight was particularly prominent in female dogs, with higher weights predicting higher activity levels and exploration during puppy tests and higher defence drive and 'hardiness' at adult tests. It was also found that changing the whelping bed substrate from corrugated cardboard to a soft bedding negatively affected puppies' resilience, and the authors suggest that the cardboard bedding may have simulated the effects of early handling.

Fox and Stelzner (1966) also investigated the effects of early experience on puppies. In this study, puppies were grouped into a control group (raised under 'typical' conditions with the mother and daily human husbandry), a handling group (handled daily for the first five weeks of life) and an isolation group (raised identically for the first four weeks of life, then single housed in a darkened room with minimal human contact for the fifth week). Handling included exposure to light, sound, changes in position and orientation, temperature, and after three weeks, human play. When tested at five weeks, handled pups showed slightly superior coordination, were more active, more social and faster in a problem solving test (although none of these difference was significant). They also showed greater initial distress at separation from human handlers, which follows what we know about attachment-type patterns in dogs (Gacsi et al., 2005).

Conversely, control pups showed greater distress when placed in a novel environment but less so in response to separation from human handlers, suggesting that attachment patterns had not formed. ECG testing also showed that handled pups had greater heart maturation than controls, as determined by greater amplitude in ECGs and on physical examination at necropsy. The isolation-reared pups were hyperalert (vigilant) and rarely rested, which prevented electroencephalograph (EEG) readings being obtained. ECG readings were instead taken while puppies were lying quietly or asleep in a darkened room and restrained in a copper-gauze box lined with foam rubber. While the conditions described in this study reflect a more extreme deprivation in rearing conditions than those experienced in laboratories or breeding facilities under current UK legislation, single-housing and handling clearly have an effect on the behaviour and cardiac maturation of puppies, which is of course vital when rearing dogs destined for use in safety assessment. Figures 2.4 and 2.5 are taken from Fox and

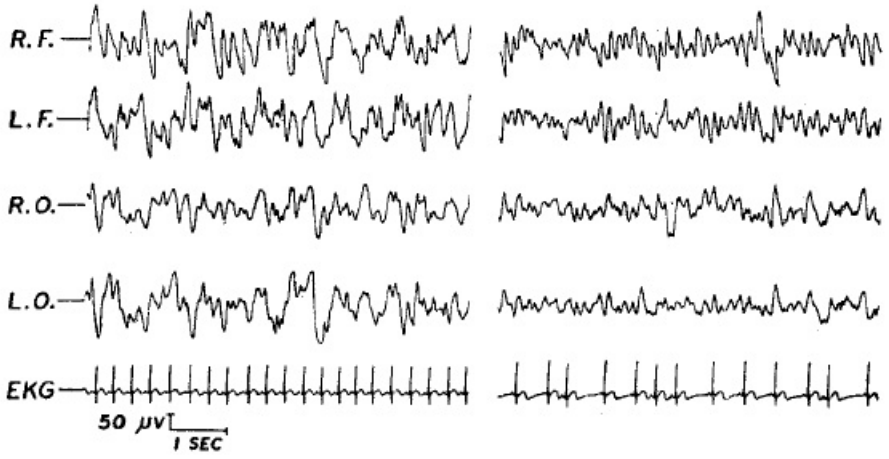


FIGURE 2.4: Adapted from Fox (1966): Sleeping EEG of handled (left) and control (right) 5-week-old pups. Greater amplitude in handled subjects is indicative of greater maturation over controls. Note bradycardia (slow heart beat) and arrhythmia (irregular heart beat) in control ECG, characteristic of normal pups of this age

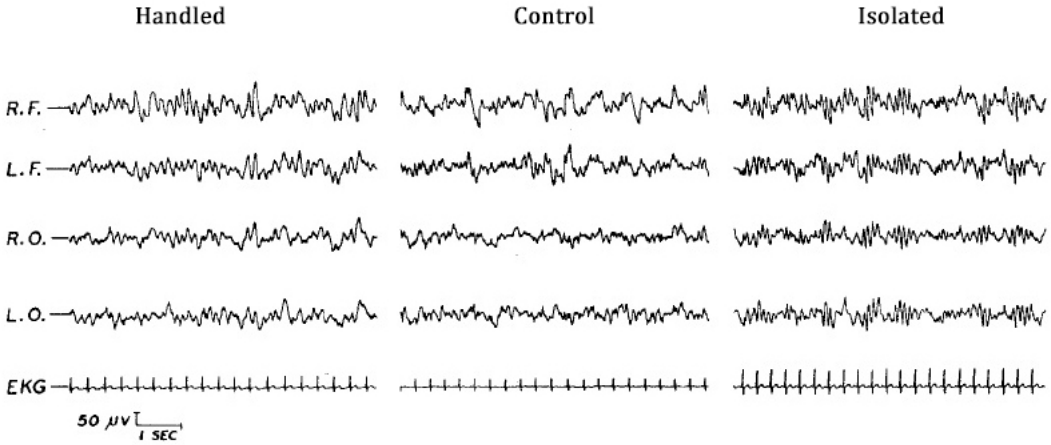


FIGURE 2.5: Adapted from Fox (1966): Awake EEG of handled (left), control (centre) and isolated (right) 5-week-old-pups. Note 'spindling' (bursts of high frequency activity) associated with extreme arousal (alerting) in socially isolated pups

Stelzner (1966) and demonstrate the ECG (EKG) and EEG patterns found across the three groups of puppies. Note the increased physical maturity and improved EEG quality demonstrated in the handled pup group, and the much poorer quality in the isolated group.

More recently, Gazzano, Mariti, Notari, Sighieri and McBride (2008) conducted an early handling programme with 43 puppies from breeding kennel and pet homes. Puppies were handled daily for the first three weeks of life and they were tested in isolation in a novel environment at eight weeks old. Puppies which had undergone handling were found to be more emotionally stable, as measured through their longer

latency to yelp, lowered duration of vocalising and the greater extent of their exploratory behaviour in the novel environment. Meunier (2006) reviewed socialisation programmes for laboratory-housed dogs and suggested that the most important factors for a 'socialisation' programme are to develop a programme which succeeds in reducing distress through the implementation of training, desensitisation and socialisation tailored to the individual future experiences of the dogs.

The importance of handling from birth is clear, as well as introducing an examination table and health checks at an early age. Where breeding takes place within the company, incorporating measures such as these into standard early rearing practices may increase the resilience of the dogs before they reach the experimental unit, where additional training for specific procedures can be introduced without the need for remedial handling. Similarly, breeding companies which supply dogs to industry must be able to provide an even greater level of resilience in dogs which are subject to additional stressors such as transport and acclimatisation.

Habituation and desensitisation to procedures and equipment is also recommended prior to the brief pre-study habituation typical of toxicology studies. Despite the wealth of literature and supporting evidence for the implementation of training, specifically positive reinforcement training, for dogs, very few examples of such research being applied in the laboratory setting exist and as such the costs and benefits to welfare are poorly understood. This is explored further in Chapter 7.

2.2.4.4 Environmental enrichment

A concern when altering the environment of a laboratory animal is that the alteration will interact with the effects of the test substance and will invalidate comparisons with historical data (Dean, 1999). However, it is important to consider enrichment of the environment when considering methods of improving the welfare of dogs housed in laboratories. When housed according to the minimum prescribed standards, dogs often lack sufficient stimuli and the space needed to display species specific behaviours (Schipper et al., 2008). Dogs kept in low-stimulus housing conditions may, for example, develop excessive fear or aggression, increased auto-grooming and vocalisations, increased passiveness, and show manipulation of enclosure barriers, repetitive locomotive behaviour (stereotypies) and corprophagy (e.g. Hetts, Derrell Clark, Calpin, Arnold & Mateo, 1992a; Hubrecht et al., 1992; Beerda, Schilder, Bernadina et al., 1999). Most of these behavioural patterns are commonly used as indicators of chronic stress (e.g. Beerda, Schilder, Bernadina et al., 1999). Schipper et al. (2008) studied 17 laboratory beagle dogs and found that the provision of a feeding enrichment

toy resulted in a decrease in inactive time and when compared to the control group, appeared to prevent an increase in the rate of stereotypies displayed over time.

There may be a reluctance to use feeding enrichment toys for laboratory-housed dogs because of a belief that it may make the dogs less willing to interact with human handlers, may encourage aggression or possessiveness within the kennel and may negatively affect health (Gaines, Rooney & Bradshaw, 2008). Gaines et al. (2008) studied this in a population of 22 military working dogs, eight of which were provided with a Kong feeding enrichment toy for four months at one hour per weekday evening. No differences were reported in working ability, nor any adverse effects on health or increase in aggression. Rather than an adverse effect on working performance, enriched dogs were found to learn faster from positive reinforcement training compared to control dogs not enriched with the feeding toy.

In a review, Wells (2004a) found that providing kennelled dogs with increased opportunities to make social contact, with conspecifics and humans, may allow the dog to gain more control over its environment, thereby decreasing the chances of the individual failing to cope with the pressures of confinement (Hubrecht et al., 1992). It is now widely recommended, based on available evidence, that kennelled dogs should be housed in pairs or groups (e.g. Hubrecht et al., 1992; Hetts, Derrell Clark, Calpin, Arnold & Mateo, 1992b; Hubrecht, 1995a; Mertens & Unshelm, 1996).
Laboratory-housed dogs have been found to show much interest in toys, particularly those which are novel in nature, can be chewed (e.g. rawhide, nylabone) or generate noise (e.g. Hubrecht & Serpell, 1993; Hubrecht, 1995a). Suspending toys slightly off the ground may also keep toys clean, increase the ease of husbandry, reduce competition between dogs and encourage interaction with the toys animals (e.g. Hubrecht & Serpell, 1993; Hubrecht, 1995b). Wells (2004a) also found that enrichment with various toys resulted in increased activity in the dogs, a desirable change in behaviour for kennelled dogs.

2.2.4.5 Impact of increased handling enrichment on health

Given the already established need for normal healthy subjects, the effects of enrichment on the brain must be taken into account. In order for laboratory animals to model normal humans, it is important that their environment supports normal development as far as possible to ensure high fidelity models.

Investigations into the effects of environmental enrichment and refinements in procedures have also found important effects on other aspects of health and behaviour,

Table 2.7: Examples of the impacts of handling on health

Study	Study animal	Findings
Nerem et al. (1980)	Rabbits	Atherosclerosis reduced in those handled in a consistent and friendly manner when compared to picking up and restraining.
Seabrook (1984)	Dairy cows	Higher yield in those handled by 'friendly' stockmen than in those handled in the normal humane method.
Meaney et al. (1988, 1995)	Rats	Handling of pups for a few minutes per day in first three weeks of life significantly decreases behavioural and endocrine responses to stress and is maintained throughout life.
Bhatnagar and Meaney (1995)	Rats	Early handling results in reduced vulnerability to stress-related diseases.
Waiblinger et al. (2004)	Cattle	Gentle handling resulted in lower heart rate and stress-related behaviours when examined by the handler
Meijer and Codogno (2006)	Mice	Frequent handling and enrichment reduced heart rate and body temperature following aversive procedures such as injection or restraint.

and as already discussed, healthy animals are key to ensuring quality of science. Table 2.7 summarises key findings.

These studies show that as well as being able to influence neural development through changes in the environment, it is possible to positively influence health in captive animals. Again, these positive changes ensure as healthy subjects as far as possible and therefore high fidelity models, as well as promoting good welfare.

Although there is some resistance to implementing enrichment because of the perceived loss of standardisation, there is increasing evidence that enrichment need not result in greater variability, and indeed that high levels of standardisation may not be beneficial to science. Würbel (2000) and Würbel and Garner (2007) argue that institutional standardisation leads to poor external validity and reduces replicability

between laboratories because of systematic differences. In line with this, they suggest that it would be better to design factorial experiments. While Benefiel et al. (2005) found that enrichments sometimes lead to increased variability, Hubrecht (2010) reviewed the literature and found that for many of the studies which have found conflicting or negative effects of enrichment, the purpose of the enrichment was not always with the aim of improving welfare. Often, more elaborate enrichment was used than was appropriate for the animals under study, or the enrichment was not necessarily appropriate for the animals. This may account for the variability in success of various enrichment strategies, as such enrichment should always aim to improve welfare, however the term may be misused. Wolfer et al. (2004) specifically investigated the effects of cage enrichment on mice. Three strains of mice were acquired by three laboratories and kept one of three housing conditions: small, standard or enriched cages. Those housed in small and standard cages showed impaired brain development, abnormal repetitive behaviours and anxious behaviours. The authors state that there was little variation in results between the laboratories or between the groups of mice when they were tested using four standard behavioural tests (an elevated O-maze, open-field test, novel-object test and place navigation in a water maze). Conversely, Crabbe, Wahlsten and Dudek (1999) investigated the effects of rigorous standardisation in mouse husbandry across three laboratories and found marked and systematic differences between laboratories despite the measures taken to ensure practices were identical.

Würbel offers an explanation for this. He suggests that since phenotypes differ depending on housing conditions, standardisation may not be optimal for good science. Standardisation produces results idiosyncratic to that particular environment and encourages systematic variation. Promoting natural behaviour and the corresponding physiological responses may be more valid. Enriched animals may actually be less sensitive to environmental idiosyncrasies (Würbel, 2001). Since standard laboratory cages impose constraints on behaviour and brain development, resulting in behavioural abnormalities and aberrant brain functions these animals may make poor models for humans. An environment which meets an animal's needs may guarantee normal behavioural and brain development (Würbel, 2002). Therefore animals which have more normal development and behaviour may be more robust to variations caused by changes in the environment as long as the environment does not cause excessive stress or inhibit the ability to perform natural behaviours. Würbel (2002) suggests that rather than attempting to eliminate the effects of environmental factors through standardisation, or to 'explain them away' by simply listing them, to promote good science it is desirable to systematically vary environmental factors and use factorial designs to investigate the effects of factors. This may reveal biologically relevant

interactions between the genetic and environmental background of the animal while reducing the limited external validity of results which is caused by high levels of standardisation.

Although there are clearly many important factors in ensuring quality in scientific research, from the design to the analysis and publication of results, welfare has a central role to play. It has been shown that changes in welfare exert influence over not just behaviour, but physiology and immunology which are of course of great interest to those using animals in scientific research. Some authors have advised against using environmental enrichment to prevent confounds through loss of standardisation. However, others have shown that improving welfare through environmental enrichment promotes more normal physiology and behaviour in animals, which is of course important when these animals are models for humans, and also may results in animals which are less sensitive to the idiosyncrasies between laboratories. As Crabbe et al. (1999) showed, even when the level of standardisation is high there is variation between laboratories resulting in unpredictable differences in the animals. To ensure good quality in science, harmonisation in welfare needs to be achieved, to allow animals a variety of coping mechanisms and to prevent decreases in welfare which have a negative impact on data.

The principles in measuring animal welfare and in improving quality of science are the central theme of this project. The purpose of the project is to identify valid measures of welfare in the laboratory dog and use these measures to empirically examine the link between welfare and quality of data output. The factors ensuring quality in science have been laid out by Poole (1997) and Festing (2010), and discussed in this chapter. The question posed: what is the link between welfare and quality of science is an important one which has not yet been addressed in the laboratory dog. What is clear is that without understanding the impact of the laboratory environment on dog welfare, there must be variables which are neither accounted for or understood in their impact on data, and so high quality science is difficult to achieve. The studies forming the majority of the research have been designed to ensure they meet the criteria required of good science and promote good scientific practice in the use of the dog.

2.3 Aims of the thesis

The welfare of laboratory-housed dogs has been studied little since Hubrecht and Beerda's studies in the 1990s. The literature tells us that emotional state, ability to cope with stressors (welfare) and life history can all have an effect on physical health. Dogs are primarily used as models of healthy humans in safety testing of new

medicines and other chemical entities and in particular the cardiovascular system. Key aspects of good scientific process and good quality of scientific data such as validity, sensitivity, reliability and repeatability are all potentially affected by varying welfare and yet there is no method of evaluating this in the dog. Although we have some behavioural, affective, and physiological measures of welfare, the relationship between these has never been studied and without understanding the internal state of the dog, easy-to-observe indicators such as behaviour cannot be said to reliably indicate welfare.

Throughout this thesis, understanding the meaning of these behavioural indicators 'for the individual' is a theme, in order to identify those most strongly associated with positive and negative welfare. Association with emotional state is a key feature, as a positive emotional state is central to positive welfare, as is the absence of suffering and the presence of positive welfare indicators.

(a) A battery of measures used in data collection is collected, across several populations of dogs in the same facility. Information about life histories and environment is used to determine which factors may influence welfare across the life cycle. The aim of this project is to identify reliable indicators of welfare. Using established measures of affect (Chapter 4) and behaviour (Chapter 5), welfare will be measured to identify reliable indicators.

(b) The literature suggests that behavioural and emotional state can influence physiology by affecting cardiac output, immune function and corticosterone response in the dog. Furthermore the link between welfare and quality of data output in other species has been made. Given the ability of dogs to suffer, the special protection offered to them and the importance of obtaining the best possible quality of data from their use, the second aim of this thesis is to determine if there is an effect of welfare on the quality of data obtained from these dogs. With assessment of cardiovascular function, a critical component of safety assessment, as well as the effects on other organ system, particular attention is paid to identifying a link between cardiovascular function and welfare. In particular, the effect of welfare on measurement error (repeatability, sensitivity, reliability) is investigated. This information is used to form a Framework for monitoring welfare and data quality.

(c) One of the barriers to implementing Refinement strategies in the laboratory environment is the ease of implementation. Behaviour is one of the most useful measures of welfare in that it can be used to provide an instantaneous measure and can be monitored with little need for training or equipment when based on a validated framework. The third aim is to identify easily-observable behavioural measures of welfare which reliably indicate changes in quality of data output and use these to measure the impact of planned Refinements. Those behavioural measures which vary

with welfare and are easy to observe in the home pen (Chapter 6) are used to form a welfare monitoring tool (Appendix E). Having identified areas in need of Refinement, the Framework and welfare monitoring tool will be used to measure the impact of a Refinement to a regulated procedure and recommendations will be made on housing and husbandry practices.

CHAPTER 3

General Methods: Developing a harmonised assessment framework for welfare and data quality

Abstract

This chapter describes the overall methodology used throughout the project. Although a number of different studies make up this thesis, a consistent methodology was employed in data collection and analysis to ensure that measures could be compared to form a coherent Framework. The methods used to measure behaviour and cardiovascular data are described here, along with the statistical analysis of data. Details of all dogs used in each study are described, as is the physical environment.

3.1 Measuring behaviour

One of the principle aims of this project was to identify behavioural indicators of welfare. The principles of using behavioural indicators of welfare require that behaviours reliably indicate a welfare state, are easily identifiable and are free from bias in interpretation. The ecological validity of behavioural measures must also be considered. In the dog, we can't consult the behavioural repertoire of wild conspecifics.

The majority of research conducted in the dog is in the pet dog. The dog-human relationship is so close that it can be difficult to disentangle behavioural responses to

the human relationship and environment and natural behaviour. One of the other focuses of research is on working dogs, which covers anything from assistance dogs, to police sniffer or patrol dogs to military working dogs.

3.2 Overview of the facility

This project was conducted in its entirety within the experimental dog unit of AstraZeneca's R&D site in the UK. A separate breeding unit on-site had been used to breed the closed colony of dogs (Alderley Park, AP, strain beagle). Like other dog units in the UK at the time, the unit had been built to replace the original dog facility, while improving dog welfare and staff efficiency. Over 200 dogs could be housed at full capacity. Each housing area was divided into three 'Zones', each with multiples rows of interlinked pens.

There were a number of design features made to the layout of dog housing areas in the unit (see Fig. 3.1 - 3.3 below) including increased visibility, interconnecting pens, indoor play areas, an outdoor play area and separate 'procedure pods' for regulated procedures. All aspects of housing and husbandry met the minimum standards set out in A(SP)A (1986) as described in Chapter 2.

3.3 Study animals

3.3.1 Development of the Welfare Assessment Framework (Chapters 4-6)

In the development of the Welfare Assessment Framework (WAF), it was necessary to identify the most suitable dogs to study, while working within the constraints off a functioning unit. Studying contrasting groups of dogs allows an investigation of factors in housing, husbandry and life history which are likely to have a positive or negative impact on welfare.

Three populations of dogs were chosen: Safety Pharmacology (SP), DMPK and Stock, because of the contrasts in factors relating to housing, husbandry and histories of regulated procedures. These differences in life history were likely to have to influenced welfare (see Chapter 2), while understanding how the chosen measures respond to differences in welfare is key to developing a framework which is sensitive. The typical use of dogs in SP and DMPK studies has been described in Chapter 2 and Table 3.1 highlights the differences between these three groups specifically within this facility.

Table 3.1: Comparison of study groups in housing, husbandry and regulated procedures

	SP	**DMPK**	**Stock**
Home pen groups	2-3	2-6	≤ 6
Care staff	One carer	Small, regular team	Varied
Regulated procedures	Regular, involving single-housing	Infrequent, brief	none

Until they were transferred to the unit, there was a common life history for all dogs. This differs from dogs which are obtained from commercial breeders, which is a much more common approach than having an internal breeding facility. Few dog facilities breed their own dogs (see Chapter 1). Typically, dogs left the breeding facility and were transferred to the unit shortly before use. A habituation protocol was applied to all puppies on a weekly basis (including a health check), which coincided with a health check, however no positive reinforcement training or desensitisation was used. Desensitisation is used to mitigate the negative impact of aversive events on welfare and quality of data output (see Chapter 7 for an example of this).

Tables 3.2 - 3.4 show the demographic details for the three groups[1], including age, time on the unit and history of company studies. The number of company studies represents a complete study, with no information on the duration, regulated procedures or severity, as this information was not available. For SP dogs, the number of studies does not include their telemetry surgery, although this does constitute a regulated procedure. The differences in the nature of SP and DMPK studies (Chapter 2 and Table 3.1 above) means that the absolute number of studies is not meaningful when comparing the two groups.

3.3.2 Refining oral gavage in the dog (Chapter 7)

Chapter 7 presents a study using a sample of 18 naive female dogs. This sample differed in several ways from the dogs presented above in the Welfare Assessment Framework. At the time of conducting the study, a limited number of dogs, all female were available for use. These dogs were also older than would be typical for a toxicology study, at less than one year old (Tables 3.5 - 3.6).

[1]Ringo and Bouncer (SP) had not undergone any regulated procedures other than telemetry surgery when data collection started, however they did undergo company studies during the data collection period. The number of company studies experienced by three DMPK dogs was not available. There had been no regulated procedures conducted on the Stock dog group.

Table 3.2: Demographic details of SP dogs

Name	ID	Time on unit (mo)	Studies	Age at start of study (mo)	Housed with
Bert	65-09	19	2	32	Ernie
Ernie	214-09	16	6	31	Bert
George	243-09	16	4	30	Bob
Bob	247-09	16	6	30	George
Peewee	275-09	15	5	29	Nibbler
Nibbler	497-09	14	4	25	Peewee
Ringo	200-10	7	0 (4)	20	Bouncer
Bouncer	207-10	7	0 (1)	20	Ringo

Table 3.3: Demographic details of DMPK dogs

Study ID	ID	Time on unit (mo)	Studies	Age at start of study (mo)	Housed with
F25	646-08	28	6	42	F26 & 1 other
F26	211-10	14	unavail.	26	F25 & 1 other
F37	289-10	14	unavail.	25	F38 & 4 others
F38	307-10	24	unavail.	26	F37 & 4 others
M2	640-09	unavail.	13	unavail.	M13
M8	623-09	20	3	33	M12 & 1 other
M12	571-09	20	5	34	M8 & 1 other
M13	131-10	13	7	27	M2

Table 3.4: Demographic details of Stock dogs

ID	Time on unit (mo)	Age at start of study (mo)	Housed with
F268-10	14	25	F273, F366 & 9 others
F273-10	7	25	F268, F366 & 9 others
F366-10	7	23	F268, F273 & 9 others
M206-10	9	26	M292 or M1 & 3 others
M292-10	7	25	M206 or M1 & 3 others
M1-11	1	18	M206 or M292 & 3 others

Table 3.5: Demographic details of Control Group dogs

Study ID	ID Number	Time on unit (mo)	Age (mo)
F1	262-11	5	24
F2	303-11	5	23
F3	309-11	5	22
F4	341-11	5	22
F5	353-11	5	21
F6	362-11	5	21

Table 3.6: Demographic details of SD Group dogs

Study ID	ID Number	Time on unit (mo)	Age (mo)
F7	204-11	5	25
F8	259-11	5	24
F9	292-11	5	23
F10	311-11	5	22
F11	343-11	5	22
F12	348-11	5	21

Table 3.7: Demographic details of RP Group dogs

Study ID	ID Number	Time on unit (mo)	Age (mo)
F13	205-11	5	25
F14	270-11	5	24
F15	312-11	5	22
F16	361-11	5	21
F17	357-11	5	21
F18	326-11	5	22

3.4 Housing and husbandry

Several areas of the dog facility were used during data collection. All baseline recording of behaviour (and cardiovascular data for SP dogs) took place in the dogs' home pens (Chapter 5). For recording of data for challenges (Chapter 6), baseline data were recorded in the home pen, while data for two of the challenges were gathered in the procedure pods or indoor play areas closest to the home pen (Table 3.8). Cognitive bias testing took place in a nearby surgery recovery suite (SP only, Table 3.9) or two adjacent indoor play areas (DMPK and Stock, Table 3.9). All data for Chapter 7 was

FIGURE 3.1: Example of housing zone (excluding SP dogs)

collected within the home pens and procedure pods of the same layout and so the dimensions are not repeated (see Section 3.4.0.1 below).

3.4.0.1 Housing

Although the three groups used in the study were drawn from separate populations, the housing for each group shared some common features. These are shown in Table 3.8.

TABLE 3.8: Description of housing areas common to all dogs

Area	Floor area (m^2)	Use
Home pen	4.84	Day-to-day housing of dogs while on-study
Indoor play areas	19.1	Daily exercise
Outdoor play area	Approx. 150 m^2	Weekly exercise

Home pens and indoor play areas were of identical size and layout throughout the unit. An example of home pen design is shown in is shown in Figure 3.1.

Home pens measured 2.15 x 2.20 m. Each pen contained three raised benches (approximately 90 x 45 cm) at 24, 31 and 58 cm heights. Pens were fronted with horizontal bars to a height of 94 cm above which was clear Perspex. Walls between pens consisted of opaque plastic to a height of 94 cm, above which was clear Perspex. Two hatches (43 x 43 cm) joined adjacent pens. Sliding hatch covers were used to single-house dogs from group housing and separate groups with a housing Zone. For

FIGURE 3.2: Layout of home pens for all dogs

FIGURE 3.3: Design of home pens for all dogs

the majority of pens, the hatch covers were of the same opaque plastic as the walls, however in SP home pens they were made from horizontal bars, to provide scent contact during prolonged periods of single-housing.

Indoor play areas were located at one or both ends of a row of pens. Each play area had a window providing natural light at a height of 1.2 m, and contained a plastic climbing frame and various toys. Two hatches with solid plastic sliding hatch covers provided access to dogs from two adjacent rows of pens.

FIGURE 3.4: Example of indoor play areas for all dogs

FIGURE 3.5: The outdoor play area

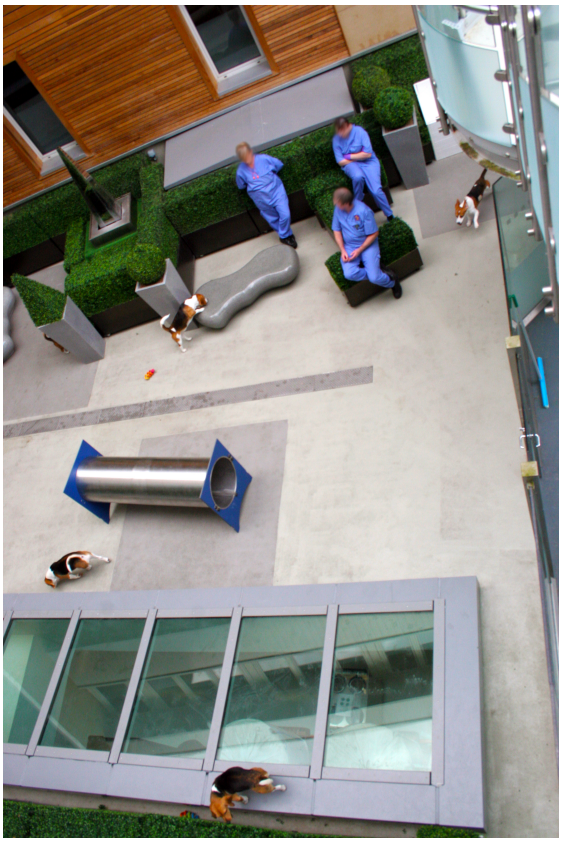


FIGURE 3.6: The outdoor play area

Procedure pods were located in each row of home pens throughout the unit, and each was a self-contained room designed to minimise the transference of excitement or agitation between dogs undergoing procedures and those in home pens. Procedure pods were also free from staff activity. These factors may not be present in other facilities and therefore the level of distress resulting from regulated procedures may vary according to the facility.

TABLE 3.9: Description of SP housing areas

Area	**Floor area (m^2)**	**Number**	**Use**
Housing zone	4.84	30	Day-to-day housing of dogs while on-study
Procedure pod	12.72	3	Health checks; regulated procedures
Indoor play area	19.1	3	Daily exercise
Outdoor play area	Approx 20	1	Daily exercise
Recovery suite	38.15	1 (5 pens)	Recovery for 24 hours post-surgery

Table 3.10: Description of DMPK and Stock housing

Area	Floor area (m^2)	Number	Use
Housing zone	4.84	36	Day-to-day housing of dogs while on-study
Procedure pod	12.72	4	Health checks; regulated procedures
Indoor play area	19.1	3	Daily exercise

3.4.1 Husbandry

Husbandry represents the most frequent occurrence of staff contact with dogs and is therefore an important consideration in welfare assessment. Husbandry practices differed in a number of ways between dogs used for different types of studies. These are detailed in the sections below.

3.4.1.1 SP

As the group of dogs most frequently on-study, the husbandry of SP dogs was largely dictated by study protocols. During daily husbandry (9am-11am), dogs in each Zone were given continual access to the three indoor and one outdoor play areas for around 30 minutes. Feeding took place between 11am-1pm, during which time dogs were single-housed in the home pen to allow accurate recording of food consumption. Final checks were conducted at around 3pm, after which time all housing Zones were locked. Staff rarely entered the dog zones outwith these times. All dogs received a once-weekly health check and weighing which took place in the procedure pod adjacent to the housing zone. SP dogs had one dedicated member of care staff who undertook all animal care duties. Regulated procedures were conducted by various licensees.

Company studies requiring telemetry recording occurred frequently in at least one of the three zones during data collection for this project. As it was not possible to obtain telemetry data from more than one dog within a single pen due to the system in use, dogs were single-housed at all times during telemetry recording.

Study protocols required that dogs were single-housed for 24 hours before the study began to obtain baseline data, 24 hours after dosing, or for the duration of the study. During recording, access to the housing zone was restricted to reduce the perceived variation in data caused by human presence. This should be contrasted with the husbandry of DMPK dogs, where factors such as single-housing and staff contact was markedly different.

Given what is known about the effects of social isolation (see Chapter 2 for details), the recognition that single-housing should be minimised in the dog (European Union, 2010), and the positive effects of human-dog interaction, it would be expected that these factors would negatively affect welfare. These patterns of daily staff activity are reflected in the pilot data used to determine activity in the dogs (Appendix C).

3.4.1.2 DMPK

The nature of the studies conducted on DMPK dogs meant that dogs spent little time on-study. The pharmacokinetic studies usually required only single doses of compounds with blood sampling taken throughout the day, with dogs being single-housed for a short time (<4hrs) after dosing. Dogs were used in rotation, which resulted in two to three months between studies for each dog. As a result husbandry was not dictated by study protocols. A small number of dedicated technicians undertook husbandry duties in DMPK. During daily husbandry (7am-11am), groups of dogs would be allowed access to the indoor and outdoor play areas. Feeding took place between 11am-1pm with all dogs group-fed, although some males (including M8 and M12) were single-housed during feeding due to competition over food. Regulated procedures were conducted by the same technicians who were responsible for daily husbandry duties. As there was frequently at least one study on-going during data collection for this project, and procedure pods were not in a separate area to the home pens, members of staff were regularly in the housing zones throughout the day. Final checks were conducted at 3pm after which housing zones were locked.

3.4.1.3 Stock

The stock dogs selected for inclusion in this project were in holding for upcoming studies. At the time of the project, dogs had come on to the unit slightly later than would be normal, however the sample chosen were the most recent arrivals as dogs were being allocated to studies on the basis of age and these would be the last dogs to be allocated. None of the dogs in the Stock group had previously been used in any study, nor undergone any regulated procedures. They had however been used as 'companion animals' during a one week discovery study: this involved them being housed alongside dogs single-housed during the study and potentially exposed to dogs experiencing the effects of regulated procedures and compounds.

The husbandry of the stock dogs was undertaken by a variety of staff. Groups housed within the same zone were frequently allowed to mix, although this was less common with male dogs due to increased levels of agonistic behaviours when unfamiliar dogs

were group housed. The groups of females in the Zone were regularly housed together in a group of 12 which was allowed access to two rows of pens and one indoor play area. As care staff were allocated a large number of stock dogs, the timing of husbandry procedures varied, either between 7am-11am, or 1pm-3pm. Dogs were moved to an indoor play area during husbandry and were allowed access to the outdoor play area once weekly when possible. As there were no frequent regulated procedures conducted in Zone 3, the only staff contact received by these dogs was during removal from the pen for husbandry and the weekly health check, although there was a high staff presence between 7am -11am due to company studies being conducted in adjacent areas. Outwith husbandry times, staff were not frequently present in this area until final checks at 3pm.

3.4.1.4 Refining oral gavage study dogs (Chapter 7)

Daily husbandry was conducted by the responsible Animal Technician on weekdays during the study. The same technician had also been responsible for the dogs for some time before the study began. This is again in contrast to facilities where dog are received from a commercial supplier a short time before a study begins. Animal care staff were responsible for husbandry at the weekends. Dogs were group-housed in their study groups (each n=6) in six interlinked pens other than for daily feeding or study protocols. Health checks, weighing and dosing all took place in the procedure pod nearest each group's home pens.

3.5 Data collection and analysis

3.5.1 Behaviour

There are a number of potential methods to measure behaviour and the most suitable is often a balance between obtaining the greatest level of accuracy and the time required to collect data. One of the first considerations for this project was the method of recording behaviour. Recording in person was not feasible because of the response of the dogs to the presence of a person. It would not have been possible to determine which behaviours were a true "baseline" and which were a response to human presence. Further, live recording of data introduces the possibility of the experimenter missing behaviours. Video recording in this instance allowed simultaneous recording of multiple dogs, desirable in a time-pressured environment, and care observation from video footage.

3.5.1.1 Sampling method

Based on the behaviours observed in pilot data, it was determined that a mixture of instantaneous sampling (for behavioural states) and all-occurrence sampling (for behavioural events) was the most appropriate, given the range of short- and long-lasting behaviours which occurred. However, instantaneous sampling has the potential to be less sensitive than continuous sampling if behavioural transitions occur between sampling points.

Using video collected for pilot data, five five-minutes samples were selected and analysed to compare continuous and instantaneous sampling. Paired-samples t-tests showed no significant differences in estimated duration of observed behaviours.

Behavioural indicators of positive or negative welfare were identified from the literature (see Chapter 2), in dogs and in other laboratory-housed species. In addition to these, behaviours not described elsewhere were identified in the pilot data. This presented a large number of behaviours for data collection in Chapter 5. Appendix B shows all behaviours contained in the coding scheme, with a description and the source of the behaviour.

3.5.2 Recording of video data

To obtain data representative of dogs' "baseline" behaviour, dogs were allowed access to two interlinked pens, with pen mates kept together. This replicates arrangements when dogs are "off-study". One camera (Sony Handycam, mounted on a tripod in front of the pen) was used to record video from each pen, which resulted in two simultaneous video streams for each sample. Videos were synchronised for analysis in The Observer XT.

A limited time was available for recording of baseline and challenge data in the morning between daily husbandry activities and feeding (roughly 7am-11am). Pilot data collection (Appendix C) showed that dogs were most active in this time period, and following feeding were less active, therefore data recording and other testing took place between 7am and 11am.

It was decided that 10 five-minute samples would provide sufficient measures of behaviour without adversely impacting the activities of staff on the unit whilst also providing more representative behaviour than five 10-minute samples.

Video recording of baseline (Chapter 5), challenge (Chapter 6) and oral gavage study (Chapter 7) data followed similar sampling protocols for all groups, however there were differences in the recording systems used. These are detailed below in Section 3.5.3.

3.5.3 Selection of samples

3.5.3.1 SP

The selection of sampling points, in particular for baseline data, was largely dictated by availability of cardiovascular data. Whilst 'on-study' dogs would normally be individually confined to one home pen. This prevents loss of signal caused by moving between pens or blocking by other dogs. During recording for baseline and challenge data dogs were housed in their 'off-study' group arrangement which resulted in significant periods of data loss during acquisition. The samples selected were largely chosen based on the availability of five minutes of continuous cardiovascular data.

Due to the time required to set up recording and the restriction on other activities on the floor, it was decided in conjunction with staff to continuously record the dogs for between 1.5 and 2 hours on each recording day, from which two five-minute samples would be selected. These samples were taken approximately 30 minutes following switching on and as far apart as possible (mean=19.62 minutes, SD=9.95) within the recording to preserve independence of observation. As the synchronisation of video and CV data was vital for planned future analysis of this study, care was taken to match the five-minute samples of video with the corresponding five-minute samples of CV data; to this end, downloaded video was trimmed into a five-minute clip for each sample.

3.5.3.2 DMPK and Stock

As only behavioural data were recorded from DMPK and Stock dogs, it was not necessary to select samples of video from longer duration videos. Company studies placed few restrictions on recording. The order in which observations of dogs were recorded on each day varied according to the activities in the rest of the housing area. Recording for individual dogs took place when there were no husbandry duties taking place in their own or immediately adjacent pens.

3.5.4 Welfare Assessment Framework

3.5.4.1 SP

The nature of Safety Pharmacology studies meant that different arrangements had to be made to obtain video data. Each home pen had a CCTV camera (Honeywell) located in the ceiling directly above it for monitoring dogs during SP studies, allowing the recording of one video stream per pen. Each recording session required significant staff input to remove pen furniture and set up telemetry recording which resulted in data collection for SP dogs taking considerably longer than for the other groups. Staff monitoring of telemetered data had resulted in a policy of allowing dogs a period of 30 minutes to 'calm down' following the disturbance caused by switching on telemetry. As a result, all samples began a minimum of 30 minutes after telemetry was switched on. It was important to ensure that behaviour measured at "baseline" (Chapter 5) or in response to a challenge (Chapter 6) was not influenced by a response to unexpected human presence.

The video stream was constantly transmitted to a remote server located in the S floor telemetry suite. Video could be downloaded from the server in one-hour blocks of Windows Media Video files with a resolution of approximately 288 x 350 pixels. The one-hour blocks of video were trimmed to provide five-minute samples as larger videos were incompatible with The Observer XT.

3.5.4.2 DMPK and Stock

The protocol for recording video was the same for both DMPK and Stock groups. One Sony Handycam camcorder was positioned on a tripod immediately in front of each of the two home pens for recording of baseline and challenge data. The camcorders were positioned such that it was possible to view all areas of the home pen, preventing data loss due to dogs moving out of sight. Unlike the areas in which SP dogs were housed, DMPK and Stock dog areas were subject to regular staff presence and it was not necessary or possible to provide a period of 'calming down' after recording began. Camcorders recorded video for 10-15 minutes for each five-minute sample to ensure that there was time for the experimenter to move out of sight.

3.5.5 Refining oral gavage study dogs (Chapter 7)

The protocol for recording video was similar to that used for DMPK and Stock dogs. Video was recorded simultaneously for all three groups using six Sony Handycam

camcorders (two per group). A camcorder was also placed in the procedure pod to record video during training, sham dosing and dosing sessions. Details of the specific sampling protocol used can be found in Chapter 7.

3.5.6 Recording of cardiovascular data (SP only)

Dogs had previously been surgically fitted with a telemetry device (Data Sciences International Chronic Use TL11M3-D70-PCTP Implant) capable of measuring ECG, femoral blood pressure and left ventricular pressure. These devices could be switched on by running a magnet across the left chest wall of the dogs.

Each home pen in the SP zone contained four telemetry receivers (Data Sciences International RMC-1) located under plates of glass in the floor. These transmitted a signal to a remote computer in an adjacent room. Notocord software was used to acquire, visualise and extract cardiovascular data which were exported to an Excel format spreadsheet.

Changes from typical housing were necessary for the acquisition of telemetered data. Telemetry signals were subject to interference from benches in the home pen and so these were removed during recording. It was only possible to record telemetered data from one dog from one home pen group at any one time as signals were also subject to interference from other dogs' signals.

3.5.7 Analysis of video data

Analysis was conducted using The Observer XT 10.5, using instantaneous (behavioural states) and all-occurrence (behavioural events) sampling. Instantaneous sampling was conducted on a 30-second interval. Appendix B shows the coding scheme used for analysis. Two video streams were available for each sample, these were synchronised within The Observer. Using the analysis tools within Observer, the duration of behavioural states and the rate per minute of behavioural events over the observed time were calculated.

3.5.8 Data analysis

All behavioural and cardiovascular data were extracted into a spread sheet using the export function of The Observer, with a total of 10 five-minute baseline and 17 five-minute challenge observations for each dog (the number and type of observation for the oral gavage study can be found in Chapter 7). Behavioural states were

presented as a proportion of observed time whilst behavioural events were presented as a rate per minute of observed time. SBP and DBP were presented as mmHg, HR as beats per minute and QTc as milliseconds.

Many of the proportional behavioural data were found not to be normally distributed. An angular transformation was performed using the formula

$$degrees(asin(\sqrt{x})) \tag{3.1}$$

where x is the original proportion. This transformation brought much of the data into normal distribution and allowed the use of parametric tests. This transformation also resulted in data being presented as percentages of total time. The rate of behavioural events was also transformed to give a rate per hour (x*60, where x is the original rate per minute), as many events occurred at less than one per minute and this allowed data to be more clearly presented. Behaviours occurring for less than 5% of time or less than twice per hour for at least 50% of subjects were excluded from analysis as these behaviours did not occur at sufficient frequency to be easily observed. Further details on the analysis performed for each study within the project can be found in the relevant chapters.

Analysis of cardiovascular data focuses on between-groups variation and within-groups variation. All systolic and diastolic blood pressure, and heart rate data were normally-distributed; analysis of variance was conducted using a one-way ANOVA at baseline (Chapter 5), and factorial ANOVA to examine the effects of challenges (Chapter 6). When examining the effects of welfare (later defined by affective state), one key measurement of interest was the level of within-groups variance as the literature suggests that reduced welfare may either increase variance in measurement (Everds et al., 2013) or decrease it, due to floor or ceiling effects (Tasker, 2012). Equality of variance between two welfare types was examined using Levene's test of equality; $p<.05$ denotes unequal variances. ANOVA is robust to unequal variances except when samples sizes differ (Glass, Peckham & Sanders, 1972), which was not an issue with this analysis.

One of the greatest impacts of variance on the 3Rs and welfare is the reduction in reliability and statistical power which leads to experiments requiring greater animal numbers to detect the effect of the compound under investigation (Chapter 2). Many studies will not be formally powered, rather relying on historical data to determine sample sizes. Increasing within-group variance (i.e. within-welfare state) reduces the

reliability of the data which in turn reduces the power of the experiment. In the field of safety assessment, not factoring in this inequality of variance when welfare has not been identified can lead to underpowered experimental design and an effect may not be detected where one is present (Type II error). Festing (2010) describes this underpowering in detail (see Chapter 2).

Samples sizes in this study are relatively small, although greater than or equal to sample sizes in many safety assessment studies (Phillips et al., 2004). Statistical significance may be difficult to obtain with small sample sizes, however it is the magnitude of the effect rather than the significance level of the difference which is of interest (Hall & Everds, 2008) : the greater the effect of welfare on the variable under investigation, the lower the power of the study design. For all cardiovascular data reported in Chapters 5 and 6, 95% confidence intervals, mean differences and the effect size of welfare type are reported, using Cohen's d (Cohen, 1992):

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s.d.} \tag{3.2}$$

with Cohen's defined effect sizes as small (≥ 0.2), medium (≥ 0.5) and large (≥ 0.8), essentially a ratio of mean difference to standard deviation.

3.6 Ethical approval

Ethical approval for each of the studies was given by the Ethics Committee in the division of Psychology, University of Stirling. With the exception of Chapter 7, no study used regulated procedures or dogs which were otherwise engaged in company studies, so local ethical approval was overseen by the Industrial Partner. The protocol for ethical approval is separately described in Chapter 7 and adhered to A(SP)A (1986). Scientific and care staff oversaw study activities to ensure that the welfare of dogs was not compromised.

CHAPTER 4

Assessing the affective state of three populations of laboratory-housed dogs

Developing the Welfare Assessment Framework I

Abstract

The measurement of the subjective, emotional experience is central to the assessment of welfare state. This study utilises existing methodology in cognitive bias testing, and a novel method in mechanical pressure threshold, to assess the affective state of the three populations of dogs. Cognitive bias testing is impractical to carry out in the laboratory environment, but the identification of other welfare measures which correspond to a bias in cognition can be used to assess the affective state in a Welfare Assessment Framework. The results of this testing identifies distinct affective states which varied across the three populations of dogs, suggesting that factors including housing, husbandry and history of regulated procedures influence welfare.

4.1 Introduction

While the nature of the experience of emotion remains disputed in non-human animals (see Chapter 2), it is apparent that there is role for emotion in the consideration of welfare as it not possible to have good welfare with a negative emotional state. The adaptive function of emotion in promoting responses to aversive stimuli (see Davidson,

1992) in particular means that in the laboratory environment it is necessary to include emotion in any assessment of welfare.

4.1.1 Aims

This study aimed to test the usefulness and validity of cognitive bias as a measure of welfare (affective state) in a group of laboratory-housed beagles. The dogs used in this study are the same as are subject to behavioural, cardiovascular and physiological measurements in Chapters 5 and 6. The findings of this study are integrated with other measurements in the proceeding chapters to identify the relationship between affective state and other measures of welfare. Associations between affective state and easy-to-observe behaviours may make it possible to identify and address negatively-valenced welfare in dogs during routine husbandry and regulated procedures, the crucial first step in addressing welfare concerns.

4.1.2 Emotion

Measuring emotion in nonverbal animals remains difficult yet central to welfare assessment. Suffering (Chapter 2), which is almost synonymous with negative emotion may be more easy to measure than positive emotion, which in many cases can be considered to be the absence of negative states. However, positive emotion includes states such as happiness and excitement, and ought to be identified and promoted in its own right (Phillips et al., 2004; Boissy et al., 2007; Yeates & Main, 2008).

The origins of the protocol used to measure 'cognitive bias' are in concepts found in human psychology. The theory of emotion biasing interpretation of ambiguous information in humans is well established. Depression influences cognitive bias at various levels of perception, including attention, interpretation, memory, cognitive control. Recent research suggests that these processes interact to bias in perception (Everaert, Koster & Derakshan, 2012). Schwarz (2000) found that anxious people are less likely to over-estimate the probability of a positive outcome in an ambiguous context than non-anxious people. There have also been theories linking the dimensions of personality (Eysenck, 1991) with an interpretational bias (Byrne & Eysenck, 1993) and showing a link between anxious personalities and the startle response (Corr et al., 1995). Since Harding, Paul and Mendl (2004) first measured a judgement 'bias' in rats similar to that present in depressed humans, there have been many studies with differing methodologies and using different species. Whilst it is necessary to adjust methodology to suit species, broadly similar methodologies have found similar effects across different species, suggesting the usefulness of cognitive bias as a cross-species

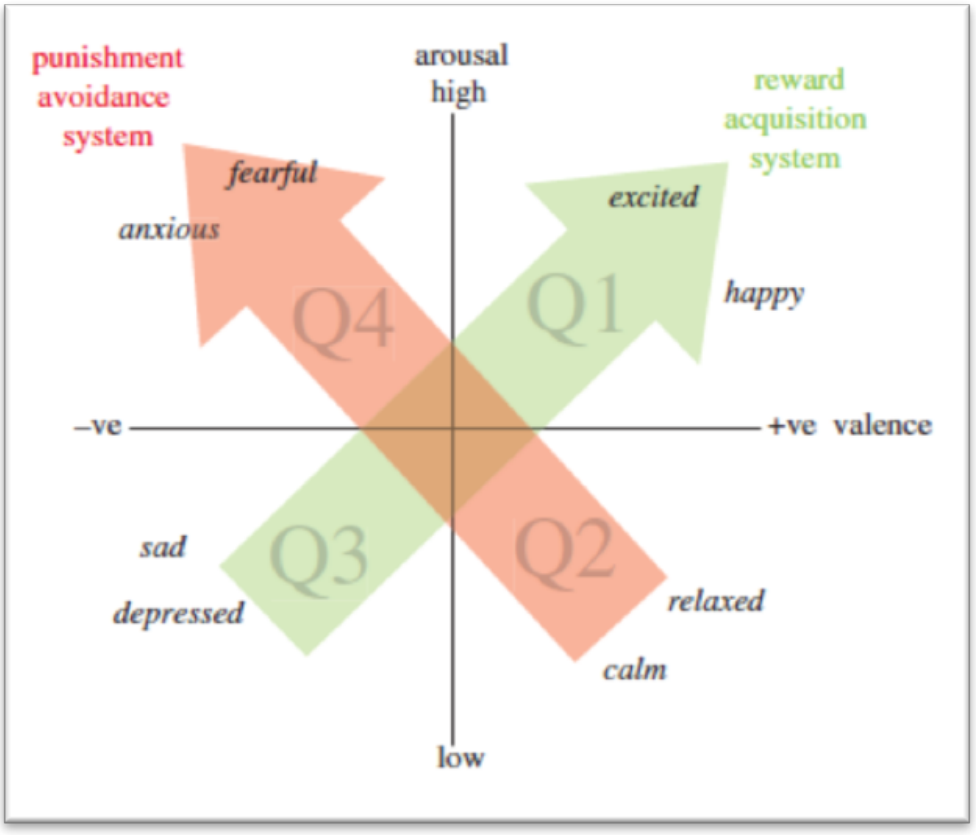


FIGURE 4.1: Adapted from Mendl et al. (2010). Core affect represented in two-dimensional space. Words in italics indicate possible locations of specific reported affective states (including discrete/basic emotions).Positive affective states are in quadrants Q1 and Q2, and negative states in quadrants Q3 and Q4. Arrows indicate putative biobehavioural systems associated with reward acquisition and the Q3Q1 axis of core affect (green), and punishment avoidance and the Q2Q4 axis of core affect (red).

measure of affective state. There is currently little integration of cognitive bias with other measures of welfare, which means that it is unclear precisely which components of judgement bias this method is detecting.

Mendl, Burman and Paul (2010) suggested a dimensional approach to understanding emotion (see Figure 4.1 for a graphical representation of this). The authors suggest that an alternative to simply measuring behavioural responses to situations believed to be positive or aversive is to develop species-specific measures of positive and negative affect using the dimensional approach and measure these in response to positive or aversive events. Responses to positive and negative events may not be the same across all animals and measuring behaviour may not elucidate the underlying meaning. Negative affect can result in active coping strategies (increased stereotypic or vigilant behaviours, for example), while prolonged and uncontrollable stress can lead to a state of learned helplessness. This level of arousal (y-axis) predicts how the animal responds

when faced with a situation in which the animal wishes to avoid a stimulus (punishment avoidance). Similarly, this also predicts how the animal will predict to a positive stimulus (reward acquisition). In low arousal, for an example an animal with a positive affective state and low drive to obtain a rewarding stimulus, little response will be seen, however an animal in a higher state of arousal will exhibit a greater behavioural response to a reward. This is consistent with the concept of the 'paranoid optimist', described by Haselton and Nettle (2006), a theory which depicts a cautious risk-taking as a successful evolutionary strategy. This relationship between arousal and affect has been demonstrated in laboratory-housed rats (van der Harst, Baars & Spruijt, 2003) in which a greater number of behavioural transitions is seen in animals following aversive events.

As an affective bias is likely to affect how an animal interprets a stimulus, different animals might respond in different ways to the same stimulus. Therefore, acute emotional responses in behaviour and physiology may vary in animals exposed to the same event, but experiencing different background mood states. These behaviours understood in the context of the dimensional approach may be a useful indicator of underlying mood.

Panksepp (2011) described seven emotional systems which are found across all mammalian species: seeking, rage, fear, lust, care, panic/grief and play. These emotions can be induced by electrical stimulation in the brain and this response appears to be common across many species. Two dimensions in particular, seeking and fear, are related to the approach/withdraw mechanism present in many species as a survival mechanism and to aid in interpretation of ambiguous, or potentially threatening stimuli. Under stress, decision-making changes to increase risk aversion, narrow focus (Blackwell et al., 2010) and become more sensitive to reward loss (Burman, Parker, Paul & Mendl, 2008). There is concern when experience has led dogs to become risk aversive, not only because it reflects a changes in their wellbeing and therefore potentially welfare state, but also because it inhibits their ability to explore the environment, learn (Blackwell et al., 2010) and through learning develop resilience to future aversive events (Lyons et al., 2009). Seery (2011) also suggests that the cardiovascular response to manageable challenges and threats is different, influenced by coping ability. Exposure to moderate stress can increase resilience, but when it becomes chronic and uncontrollable, learning and welfare are negatively affected. Differentiating between these two types of response to stress and the effect on emotional state is crucial to welfare assessment.

4.1.3 Use of cognitive bias testing in other species

TABLE 4.1: Species in which cognitive bias methodologies have been employed

Species	Publication	Methodology	Findings
Rats	Harding et al. (2004)	Go/no go lever press	Unpredictable housing induced -ve bias
	Brydges, Leach, Nicol, Wright and Bateson (2011)	Go/go lever press	Enriched cages induced +ve bias
	Rygula, Pluta and Popik (2012)	Go/go lever press	Human tickling induced +ve bias
Mice	Boleij et al. (2012)	Go/no go odour cue	Aversive bright light induced -ve bias
Capuchins	Pomerantz, Terkel, Suomi and Paukner (2012)	Go/no go	Stereotypies associated with -ve bias
Macaques	Bethell, Holmes, Maclarnon and Semple (2012)	Go/no go visual cue	Health check induced -ve bias
Grizzly bears	Keen et al. (2013)	Paw/muzzle touch	No difference following 2 hours with enrichment items
Holstein calves	Neave, Daros, Costa, von Keyserlingk and Weary (2013)	Go/no go visual cue	Dehorning induced -ve bias
Chicks	Salmeto et al. (2011)	Runaway response from threat	Isolation induced -ve bias
Pigs	Douglas, Bateson, Walsh, Bédué and Edwards (2012)	Go/no go auditory cue	Enrichment induced +ve bias
Sheep	Doyle, Fisher, Hinch, Boissy and Lee (2010)	Go/no go location	Release from restraint induced +ve bias
	Doyle, Hinch et al. (2011)	Go/no go location	Serotonin inhibitor induces -ve bias
Goats	Briefer and McElligott (2013)	Go/no go location	Females exhibited more +ve bias several years following neglect; males did not
Bees	Bateson, Desire, Gartside and Wright (2011)	Go/no go	-ve bias in agitated bees
Starlings	Matheson et al. (2008)	Go/no go light cue	Enriched cages induced +ve bias
	Brilot, Asher and Bateson (2010)	Go/no go shade	Stereotyping associated with -ve bias
Dogs	Mendl, Brooks et al. (2010)	G/no go shade cue	Separation behaviours associated with -ve bias
	Burman et al. (2011)	Go/no go shade cue	Food reward induced +ve bias
	Müller et al. (2012)	Go/no go location	Owner absence did not cause pessimistic bias
Review	Mendl, Burman, Parker and Paul (2009)		

Cognitive bias testing has been employed in many species since Harding and colleagues first tested their paradigm in rats in 2005. (Table 4.1). Most extensively used in rodents, a variety of other species have since been tested, from farm animals to nonhuman primates to zoo animals and cognitive bias testing in dogs is discussed below. The most frequently used used approach is a go/no go methodology, in which the animal is trained (through operant conditioning) to approach one stimulus and not to approach another. Latencies to approach ambiguous, intermediate stimuli or the frequency of approaches are then compared to determine a pattern.

The consistent ability of the cognitive bias paradigm (adapted to species) to detect changing biases increases confidence that it is measuring a stable trait in response to ambiguous stimuli. There is some difference in the findings of studies, with some finding that enriched housing induced a positive bias (e.g. Brydges et al., 2011) while others, found that brief enrichments had no effect (e.g. Keen et al., 2013). However, aversive events such as restraint (Doyle, Lee et al., 2011) or health checks (Bethell et al., 2012) appear to be capable of inducing negative biases. The prevailing methodology is a go/no go paradigm in which the animal chooses to respond, or not respond, based on its interpretation of ambiguous stimuli. Other paradigms require active responses (e.g. lever press) to avoid an aversive event (Rygula et al., 2012). The use of negative reinforcement is however less ethically sound in a population of vulnerable animals, and response to the unrewarded stimulus has the potential to act as a punisher, decreasing any tentative searching behaviour. Those using a go/go methodology have found the paradigm to be less sensitive, so this methodology is not suitable. The predominant drawback of the cognitive bias methodology is the length of time required to train and test the animals, in conjunction with the need for somewhat specialised equipment. This also precludes it from being used as an 'instantaneous' measure of welfare, in the way that behaviour might. This is a weakness of existing cognitive bias methodology. Integrating cognitive bias with behavioural measures, as in this project, increases its usefulness as a welfare measure.

Many of the previous studies utilising cognitive bias have done so to investigate a change in affective state following a positive or negative event, for example a change in housing. Others, such as Pomerantz et al. (2012) have investigated an underlying affective state without experimental manipulations. It is reasonable to assume that a cognitive bias showing risk aversion in an individual suggests that previous events in the animal's life have caused it to adapt a risk-averse strategy. In this population of dogs, the goal was to investigate the influence of life history, housing practices and husbandry practices on welfare and so the most suitable paradigm was to use a go/no go response type task with no experimental manipulation. In contrast to other uses of cognitive bias testing, this project examines the background affective state. It also

avoids some of the problems associated with the methodology by associating detected biases with other measures of welfare which can be more readily identified.

While a positive bias could be considered to be influenced by motivation to search for food, especially using a go/no-go paradigm with a food reward, it seems this might not be the case. Verbeek, Ferguson and Lee (2014) investigated the effects of ghrelin administration (used to induce hunger) in nine sheep and found that ghrelin induced a pessimistic bias, with sheep being less willing to search when ghrelin had been administered. This suggests that motivation for food alone is not sufficient to induce a positive bias in searching. Verbeek et al. (2014) did however find that the bias was more positive when testing did not closely follow feeding in another experiment, so conducting testing at the same time of day should ensure that hunger does not bias results.

There is a suggestion that the apparent detection of cognitive biases in bees (see Bateson et al., 2011) is sufficient to discredit the paradigm, as the mechanism in bees can be explained by simpler mechanisms such as attentional bias without emotional processing (Mendl, Paul & Chittka, 2011). However, there is evidence that invertebrates are capable of more complex cognition than previously thought (Giurfa, 2013) and a generalised stress response has been proposed in the honey bee(Even, Devaud & Barron, 2012), via a system similar to the mammalian HPA-axis which attenuates stress and fight or flight response (this may be unsurprising given its social structure) and patterns of laterality similar to those described in other species have been discovered (Rogers, Rigosi, Frasnelli & Vallortigara, 2013). Indeed, Horvath, Angeletti, Nascetti and Carere (2013) propose that invertebrate welfare is significantly under-represented in light of recent evidence of their cognitive capacities. The mechanism of cognitive bias testing should therefore not be discounted on this basis alone. It is widely understood that not even humans can always identify or understand the emotions that they feel and so conscious awareness of emotion should not be the central measure of the subjective experience.

Whether cognitive bias truly measures affective state in the sense in which emotion is understood by humans, or whether it represents an unconscious response designed to promote a survival response (i.e. attachment to kin, offspring, flight from danger, shorthand for responding) is not strictly relevant to the sense in which it is used here. It is not for this project to determine the exact nature or mechanism behind cognitive bias. For a negative cognitive bias to be associated with increased nociception and a behavioural demonstration of an inability to cope with the environment tells us enough that we do not need to understand the subjective experience. Mendl et al. (2009) reviewed previous methodologies and made suggestions based on the most successful

and valid approaches. The authors suggest that for testing to be valid, cues should be counterbalanced across positive and negative reinforcers, as where cues are close together testing may be evaluating perceptual rather than affective processes. Similarly, simultaneous presentation may measure attention bias rather than affective bias while a single probe is more likely to measure evaluation of the probe. For this reason, the visual stimuli were the same as used in Mendl, Brooks et al. (2010). Studies which have used three probes have shown differences in responses to the probes which would have been missed if only one probe had been used. This suggests that the use of multiple probes is necessary to accurately measure bias. The authors also suggest that studies which use unreinforced probes are more successful in detecting differences.

The authors also highlight the importance of continuing research identifying more 'rapid and practicable tests for use in the field' and the effects of affective biases on ability to cope and welfare in animals. One of the aims of following studies is to integrate cognitive bias testing with other measures of welfare, and identify associations and relationships. Where easily identified measures can be reliably said to predict a positive or negative affective state, these can be used to rapidly assess the emotional and welfare state of a dog.

4.1.4 The use of cognitive bias testing in dogs

In the first use of cognitive bias testing in dogs, Mendl, Brooks et al. (2010) tested a group of 24 shelter-housed dogs for separation anxiety behaviours and then for an affective bias. The test used was a location discrimination, with dogs trained using a rewarded and unrewarded location, and a further three intermediate locations were used in testing. Dogs were tested on three occasions and all dogs showed increasing latencies from rewarded to unrewarded stimuli, however dogs identified as showing separation-related behaviour (using a previously validated separation anxiety measurement tool) had significantly higher latencies to the middle and near-negative locations. Importantly, analysis detected no extinction of responses to unrewarded ambiguous stimuli across testing, suggesting that the dogs did not learn that these were always unrewarded.

In a follow-up study, used a visual discrimination task to investigate the cognitive bias of a group of dogs in which 'optimism' or 'pessimism' had been induced using a search maze which was either rewarded or unrewarded (dogs had previously been conditioned to expect a reward for searching), and a significant difference was found in latencies to ambiguous probes with the unrewarded maze resulting in higher latencies. Burman and colleagues' design was chosen for this study as it had been shown to be sensitive to

a change in cognitive bias in dogs. It also appears that in the visual discrimination task, dogs could clearly identify and distinguish between the shades used for the ambiguous probes, and the lack of a punisher for approaching the unrewarded box meant that it was appropriate for this population.

4.1.5 Nociceptive threshold and affective state

Affective state is known to influence a number of cognitive functions and responses, for example interpretation of information and the perception of pain, nociception. Pain can be described as the link between increased nociceptive threshold and depression or anxiety is well established in humans (e.g. Villemure & Bushnell, 2002; Klauenberg et al., 2008), as is the corresponding difference in affective state as measured through cognitive bias (Pincus & Morley, 2001). The existence of the same relationships in several animal species such as rodents and fish (e.g. Braithwaite & Boulcott, 2007; Wilson, Boyette-Davis & Fuchs, 2007) is becoming apparent, although Shi, Qi, Gao, Wang and Luo (2010) found an opposite pattern, with negative emotion decreasing sensitivity to pain. However a link between depression and a loss of sensitivity to pain is documented in some humans (Dickens, McGowan & Dale, 2003), suggesting a complex relationship with emotion, and perhaps representing in animals the difference between active coping and learned helplessness.

While the link between a negative affective state and nociception has implications for all captive animals, it is particularly important for those experiencing potentially painful procedures. Regulated procedures are defined as those with the potential to cause pain, suffering, distress or lasting harm (A(SP)A, 1986) and are regularly conducted as part of *in vivo* research. Especially in safety assessment research where procedures such as blood collection or dosing may also be accompanied by unpleasant side effects, the understanding of the experience of pain is crucial to understanding the impact on welfare, as reflected in the retrospective review mandated by the recent European Directive.

If a negative affective state decreases nociceptive threshold, a population of individuals with negative affective states are at greater risk of experiencing adverse effects from regulated procedures. Clearly, it is both ethically and scientifically unacceptable to have a population of study animals at greater impact of experiencing aversive effects, as this leads to a greater welfare risk and an increase in the potential for increased variance to influence the data obtained from those individuals. As the dogs in this project were already at risk of experiencing regular aversive procedures, it would not have been acceptable to subject them to a test of nociception which required the

induction of pain. For this reason it was necessary to use an analogue of nociceptive threshold, mechanical pressure threshold (MPT). A device designed to measure post-operative pain in domestic species was utilised, the Prod (Pressure rate onset device) TopCat Metrology (Dixon, Taylor, Steagall, Brondani & Luna, 2007; Dixon et al., 2010). While designed to measure pain perception (mechanical nociceptive threshold, MNT) in a wound site, in healthy, unrestrained and unsedated animals the device is capable of measuring mechanical pressure sensitivity rather than nociceptive threshold as the animals are free to move away from the device before it causes discomfort.

The MNT has been shown to significantly increase following analgesia and decrease following surgery in several species (e.g. Hunt, Grint, Murrell & Taylor, 2010; Bortolami, Murrell & Slingsby, 2013) and has been used to monitor the need for analgesia. There are two predominant advantages to using MNT as a measurement of sensitivity to mechanical pressure sensitivity over finger palpation: measuring the existence or progression of a painful condition through quantification of the intensity of pain or discomfort; the device also has rigorously applied limits beyond which no further potentially damaging or excessively painful stimuli can be applied (Jolliffe et al., 2009). These factors were particularly crucial for this study, due to the need to quantify threshold without inducing pain in the subjects.

A baseline pressure threshold for dogs was established at 5.5 ± 1.4N (Hoffmann, Kastner & Kramer, 2010), in six healthy adult beagles using an 8mm tip, although the precise methodology is not described. Hunt et al. (2010) used the Prod to measure wound sensitivity (described as mechanical nociceptive threshold, MNT) in 37 healthy dogs undergoing elective surgeries. An 8mm tip was used on the device, applied 1cm from the wound edge until the animal responded in a manner indicating pain (flinching, growling, escape attempts), with baseline (pre-surgery) MNT readings ranged between 10-15N. This provided a range of potentially 'normal' values within which to work, and a validation of the methodology in analgesia trials.

To understand and compare mechanical threshold sensitivity, it is important to take into account the tip size used on the measurement device, as the threshold varies with the size of the tip used, with the largest tips requiring the greatest force to approach the nociceptive threshold. Therefore, comparable thresholds are only possible where the area on which pressure applied is equal or force is adjusted for area of applied pressure, as:

$$Pressure = \frac{Force}{Area} \tag{4.1}$$

The inclusion of mechanical pressure threshold testing (MPTT) alongside cognitive bias testing provides a further measure of affective state, as previous research in varied species suggests that a negative affective state should cause a corresponding decrease in nociceptive threshold and therefore in MPT. Those individuals displaying a negative affective state as measured through cognitive bias testing were hypothesised to also display a lower MPT than those displaying a positive affective state. While behavioural and cardiovascular observation and testing (as described in the relevant chapters) provides information on the impact of welfare on the individual and between groups of dogs, and suggest links to the impact on quality of scientific output, for a subset of animals to experience heightened nociception as the result of poor welfare is scientifically and ethically unacceptable and it is therefore important to determine if this link suggested by literature exists in this population of dogs.

4.2 Methodology for cognitive bias testing

4.2.1 Training and testing area

SP Dogs

The area used in the study was at the time used as the recovery suite for surgical procedures, however as the number of these procedures was low, use was infrequent. The room had five standard sized pens, with a corridor between the pens. A door at one end of the room lead to the telemetry suite, at the other to an external corridor and a side door lead to the surgical area (Figure 4.2). The corridor between the pens (2) was used as the immediate area for testing, with the stimulus boxes placed 3.6m from the dog's starting point. The placing of the box mid-way along the corridor was designed to encourage dogs to explore the box; in piloting it was found that if the box was placed at the end of the room dogs were unlikely to get close enough to the box to discover the reward. The dogs' starting location was in the pen closest to the exit to the corridor (1). Whilst noise from the adjacent corridor leading from the main landing occasionally disturbed training and testing, this was determined to be the quietest location within the room.

The experimenter left the testing room into the adjacent surgical area to bait the stimulus boxes (3). This ensured that the dog was unable to see the boxes being baited and was unable to see the box until released from the starting point.

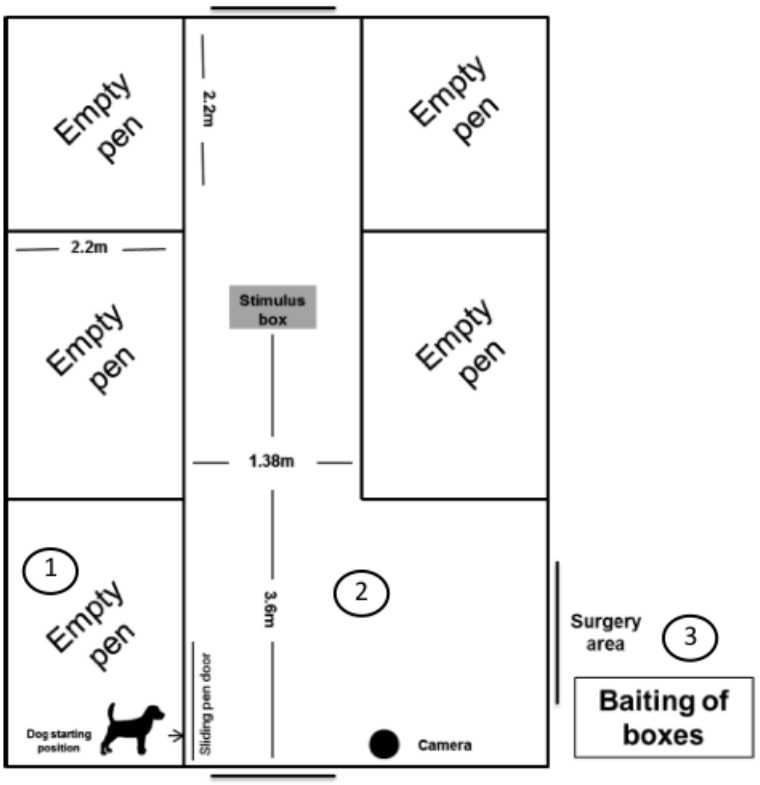


FIGURE 4.2: Layout of the testing are (SP only)

DMPK and Stock Dogs

It was necessary to find another area for training and testing for DMPK and Stock dogs. At the time of training for DMPK dogs, a row of interlinking play areas became available, each of these play areas provided a large area free of objects other than the climbing frame in each, which could be pushed aside (see Figure 4.3 for layout). The dog could be restricted to one pen (1) and released via a sliding door for each trial. The dog was unable to see the experimenter when restricted to one pen. The experimenter baited the box in the third pen (3) and the probe box was placed in the middle (2) pen for each trial. The stimulus box was placed centrally and 3.5m from the dog's starting point at the sliding door between pens 1 and 2. The camera was placed centrally behind the dog in pen 1.

DMPK dogs were located on SB floor, two floors below the testing area. Dogs were transported via a lift to the testing area. During familiarization, dogs were carried by the experimenter but quickly acclimatized to lead walking and as they appeared to find training and testing enjoyable, the anticipation of participating appeared to help them overcome anxiety caused by being introduced to an unfamiliar area. Stock dogs

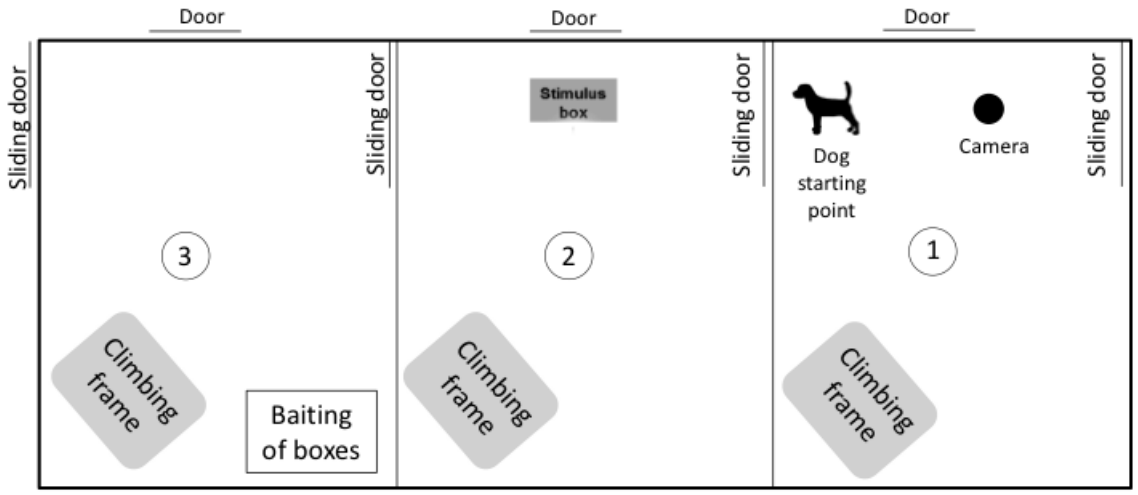


FIGURE 4.3: Layout of the testing are (DMPK and Stock)

were located on the same floor and walked from their home pens to the training and testing area that was nearby.

4.2.2 Materials

4.2.2.1 Stimulus boxes

Five cardboard boxes measuring 30 x 21 x 30 cm were used as the rewarded, unrewarded and probe stimulus boxes. A stimulus card (A4 sheet with the appropriate shade of grey) was glued to the front, bottom and back of each box. The plastic bowl (see below) was placed into the box. The bowl was placed on top of a small cardboard box approximately 4cm in height within the box. Due to their small stature, dogs were unable to reach into the box to get the food reward unless it was raised. The box was open on top and the dogs were able to search for food by searching the inside of the box. Dogs were found to be unwilling to knock the box over to obtain food. The bowl was not visible outside the box.

4.2.2.2 Stimulus cards

The shades of grey used on the stimulus cards were selected from the 'red green blue (RGB) colour model' which ranges from 0-255, with 0 being darkest. The following shades were used in this study, as in Burman et al. (2011):

- RGB 240 (greyscale 6% - very light grey). Rewarded stimulus (R).
- RGB 185 (greyscale 27.5%). Near-rewarded probe stimulus (NR).
- RGB 130 (greyscale 49%). Middle probe stimulus (M).
- RGB 75 (greyscale 70.5%). Near-unrewarded stimulus (NU).
- RGB 20 (greyscale 92% very dark grey). Unrewarded stimulus (UR).

The order of shades used as stimuli was counter-balanced between the dogs, with one half of the dogs having the very light grey box as the rewarded (R) stimulus and the other half having the very dark grey box as the R stimulus. There was a difference of 55 RGB (greyscale 21.5%) between each of the above shades. Dogs have been able to discriminate much closer shades of grey, for example in a study by Pretterer et al. (2004) which used 30 shades of grey between 0-255. It was determined based on this finding, and the success of Burman and colleague's use of the same shades, that there should be no experimental confound caused by the dogs being unable to discriminate between the shades of grey. As variations in shade can be caused by changes in printer toner across time, all of stimulus cards were printed in a single batch on one day.

4.2.2.3 Bowl

One clear plastic bowl was used to contain the food reward within the stimulus boxes during training and testing. When in the rewarded box, there was food in the bowl. When in the unrewarded box, there was no food in the bowl. To prevent odour cues, the bowl was baited with the food reward on each unrewarded trial (by placing the food in the bowl and then removing). As the same bowl was used throughout training and testing, there was no reliable odour cue. That the dogs continued to approach the stimulus boxes when no food was present in testing suggests that they were not relying on an odour cue to determine whether or not food was present.

4.2.2.4 Food reward

SP

Pedigree®'Meaty Loaf With Real Beef' diet was used as the food reward. Approximately one teaspoon was used in each trial. The diet is highly palatable and normally reserved for inappetant dogs, making it suitable for use as a food reward.

Other diets were trialled as rewards on the advice of senior staff, including SDS Dog-D3 (E) SQC (standard adult dry diet) and SDS Puppy Diet C (standard puppy dry diet), both supplied by Special Diet Services Ltd, England. In piloting and early familiarisation periods (see Section 2.5 below) these diets were found to be unattractive to dogs and were deemed unsuitable as food rewards.

DMPK and Stock

The food item used for SP dogs was highly palatable but inconvenient to use as a reward due to its soft nature. Permission was obtained from company management to obtain additional food items for use in this study. Pedigree®'Schmakos' and 'Rodeo chewy twists' were trialled with dogs and found to be highly palatable and easily consumed by both DMPK and Stock dogs. As there were differences in preference between dogs, a small amount of each was presented on each trial, although overall preference appeared to be for 'Schmakos'. Two pieces of approximately 2cm of each item was presented on each trial (approximately 8g total).

4.2.2.5 Recording equipment

All training and testing sessions were recorded using a Sony Handycam digital video camera mounted on a tripod. This allowed for accurate recording of the measures used in this study, as precise recording of response latencies was key.

4.2.3 Familiarisation, training and testing for cognitive bias testing

4.2.3.1 Familiarisation

SP

Prior to training and testing, dogs were removed from the home pen and taken to the study area by the experimenter, who was unfamiliar to the dogs at this stage. Familiarisation sessions were given up to once daily before the start of training to allow the dogs to become familiar with the experimenter, the room and unfamiliar equipment. The dogs were unused to being taken outwith their normal immediate

environment, therefore there was a potential for the novel set-up to induce anxiety that could influence results.

On the first day of familiarisation, dogs were taken to the study area in pairs with their pen mate to reduce the anxiety induced by an unfamiliar environment. Dogs were allowed to freely explore the area and become familiar with the experimenter, with no interaction being forced. If dogs appeared to be comfortable with this after two sessions, they moved on to individual familiarisation periods. Details of the sessions received by each dog can be found in Tables 4.2 - 4.4. Each session lasted around 10 minutes.

Food was presented in the clear plastic bowl in at least one familiarisation period for each dog. In piloting, it was found that while the food reward was clearly attractive, the unfamiliar bowl and the noises caused by contacting it prevented the dogs from obtaining a reward. To prevent fear of the bowl from confounding results, it was decided to place a piece of food in the bowl which was then placed on the floor for dogs to explore and eat. The number of sessions required by dogs to become comfortable with this varied.

TABLE 4.2: The number and type of familiarisation sessions received by each SP dog

Dog	**Pair sessions**	**Indv. sessions**	**Reward presented**	**Proceeded**
Bert	2	3	1	Yes
Ernie	4	4	2	No
George	2	2	1	Yes
Bob	2	2	1	Yes
Peewee	3	2	2	Yes
Nibbler	4	4	4	No
Bouncer	4	4	4	Yes
Ringo	2	4	2	Yes

Two dogs (Bouncer and Ringo) were added to the initial sample size of six as two dogs in the sample (Ernie and Nibbler) found the novel environment and experimenter aversive and were reluctant to take the food reward from the bowl on the floor. Despite additional familiarisation sessions, these dogs were unable to approach the food reward to consume it and became increasingly anxious in the unfamiliar environment. As it became apparent that training these dogs would potentially induce distress and was unlikely to succeed, they were removed from the study. These two dogs had such a negative interpretation of the unfamiliar surroundings that they were unable to participate, suggesting that they have an extremely 'pessimistic' bias. Worth noting is that of the dogs which did proceed to training, two initially appeared nervous

of the experimenter and equipment (Bert and Peewee), one appeared to adjust fairly quickly (Bob) and one seemed unfazed (George).

DMPK

Given the difficulties encountered when attempting familiarization with SP dogs, modifications to the design were made when beginning familiarization with the DMPK dogs. Cognitive bias testing had been the first study conducted with SP dogs, with behavioural observation occurring later, meaning I was unfamiliar to the dogs at the outset of familiarization. In addition to the presence of an unfamiliar person (unfamiliar people were often licensees present only for regulated procedures), dogs were required to adjust to an unfamiliar location and equipment. These factors may have made it more difficult for dogs to become relaxed enough to search for and obtain the food reward.

Table 4.3: The number and type of familiarisation sessions received by each DMPK dog

Dog	Pair sessions	Indv. sessions	Reward presented	Proceeded
F25	1	1	1	Yes
F26	1	1	1	Yes
F37	2	3	4	Yes
F38	2	4	5	No
M2	3	4	3	No
M8	1	4	4	No
M12	1	4	3	Yes
M13	0	2	2	Yes

Stock

Table 4.4: The number and type of familiarisation sessions received by each Stock dog

Dog	Pair sessions	Indv. sessions	Reward presented	Proceeded
F268	1	1	2	Yes
F273	1	1	2	Yes
F366	1	1	2	Yes
M206	1	1	2	Yes
M292	1	2	2	Yes
M1	1	1	2	Yes

4.2.3.2 Training

All dogs received training period of up to half an hour and up to once per day until a pre-determined criterion was reached. Training ended when the dog reached criterion.

The criterion to be reached specified that the dog must run faster to the rewarded (R) box than the unrewarded (UR) box for six consecutive trials and by at least 0.5 seconds. Half of these were rewarded, half unrewarded, to ensure that the dog has not simply following an olfactory cue. This criterion ensured that the dogs had reliably learned that the R box was always rewarded and the UR box was always unrewarded. As dogs approached the criterion, it was noted that there was a clear difference in their running style towards the R and UR boxes, with a distinctly rapid gait and direct movement towards the R box, and a slower gait and exploration of surroundings exhibited when the UR box was presented.

During training, the rewarded and unrewarded boxes were presented in a pre-determined random order to the dog. The order was determined by a random number generator. The first four trials of any training session always followed the pattern *'R R UR UR'*. Following Burman et al. (2011)'s reasoning, this was used to prevent extinction of a searching response caused by the UR box being presented first, and to refresh the association between stimulus cards and reward at the outset of a training session. The extinction of a searching response in training had to be carefully considered in this sample of dogs, as the boxes were unfamiliar and apparently initially aversive to the dogs. An initially unrewarded search may have extinguished a tentative searching response.

Before each trial, the dog was placed in the holding area. The experimenter left the room and baited the stimulus box. On return, the experimenter placed the box on the floor, the location of which was marked. The dog was released and given up to 30 seconds to find the food in the rewarded box or give up. Training sessions were provided up to once daily until criterion was reached. On the day before testing, dogs received a 'refresher' session to ensure that they could reach criterion within 15 trials, as this was the number of trials used in testing. Further training would be given if the dog failed this, however no dog which had reached criterion subsequently failed in the refresher session. Timing began when the dog was released from the starting position, 3m from the box. If the dog failed to reach the criterion in the first training session, training was repeated once daily until the criterion was reached.

SP

There was variation between dogs in the number of trials needed to reach criterion. The number of training trials received by each dog is shown in Table 4.5. One dog which progressed to training, Bouncer, appeared to find the training protocol aversive, and was unwilling to explore the stimulus boxes. Despite receiving two training sessions with around 20 trials in each, contact was made with the box on fewer than five occasions. As this dog was unable to overcome the neophobia induced by the

Table 4.5: The number of training trials received by each SP dog

Dog	Number of training trials	Proceeded to testing
Bert	34	Yes
George	41	Yes
Bob	42	Yes
Peewee	41	Yes
Bouncer	40	No
Ringo	65	Yes

protocol and materials, it was decided that it was not possible to train him and so he was removed from the study.

DMPK

The number of trials required by DMPK dogs varied considerably, but was overall lower than that needed by SP dogs. One dog in particular, F25, appeared to learn the task very quickly, only approaching the UR box on its first appearance in testing session 1, after which the UR was not approached or only approach slowly. F26 also appeared to find the training protocol and particularly the food item highly rewarding and she too completed training quickly. The other DMPK dogs showed a similar response, with the exception of M12, who appeared to find the presence of the experimenter stressful, but readily took part in the protocol.

Table 4.6: The number of training trials received by each DMPK dog

Dog	Number of training trials	Proceeded to testing
F25	15	Yes
F26	19	Yes
F37	47	Yes
M12	25	Yes
M13	28	Yes

Stock

Of the three groups, Stock dogs adjusted to the training protocol with the greatest ease, taking the food item via hand feeding and the from stimulus boxes in the familiarisation period and showing no signs of distress as a result of being taken to the testing area. This resulted in fewer sessions being required to reach the criterion. The dogs also showed fewer behavioural signs of excitement or nervousness during the training sessions (this is also reflected in Chapters 5 and 6). F268 was unable to reach into the boxes so the food bowl was raised by approximately 2cm, while M292 showed some signs of being vigilant and lower body posture but was no less willing to participate. F273 and F366 quickly returned to the starting position after each trial

without having to be lifted while M206 and M1 showed some signs of excitement but this did not interfere with the training protocol as it had done with dogs in previous groups.

TABLE 4.7: The number and type of training trials received by each Stock dog

Dog	Number of training trials	Proceeded to testing
F268	26	Yes
F273	32	Yes
F366	31	Yes
M206	24	Yes
M292	31	Yes
M1	18	Yes

4.2.3.3 Testing

During the testing phase, dogs were exposed to a sequence of 15 trials of pre-determined order. In order to minimize chances of dogs realising that the probe boxes were always unrewarded (again, to prevent olfactory cues), presentation of probe boxes was interspersed with presentation of the rewarded and unrewarded boxes. In order to present each probe box six times, all dogs received six testing sessions, one per day over two consecutive weeks. The order of the probe boxes was counterbalanced within each trial, however the order of R(+) and UR(-) boxes remained the same:

$$+--+M+-+NR-++-NU$$

4.3 Methodology for Mechanical Pressure Threshold Testing

4.3.1 Subjects

MPTT took place after the completion of other studies comprising the Welfare Assessment Framework, once the necessary equipment had been obtained, and the dogs tested had previously been subject to behavioural, cardiovascular and cognitive bias measurements in its development. Testing took place some time after the conclusion of the other studies comprising the welfare assessment framework and therefore the number of dogs available for use in MPTT was smaller. Some dogs had been euthanised prior to MPTT. Due to the difference in time between completing data collection for other parts of the welfare assessment framework (SP, eight months;

DMPK, five months; Stock, three months), it was important to determine if the remaining dogs' welfare had changed over time. If so, previously obtained measures of cognitive bias would no longer be valid and could not be compared to measures obtained during MPTT, without conducting cognitive bias testing once more, which was not feasible at the time of the study. To this end, dogs were observed for five minutes using the measures and methodology used throughout behavioural data collection. This was compared to previously observed baseline behaviour for behaviours determined to be indicative of good or poor welfare (see Chapters 5 and 6). Only one dog (Stock M206) was found to be exhibiting a change in welfare, as seen in an increased in rapid locomotion, alert behaviour and stereotypies (e.g. paw lifts, lip smacking). This is likely to be due to his single-housing for a period of three weeks as a 'companion' to another singly-housed dog undergoing a company study immediately before MPTT took place. However as there was limited time to test dogs, behaviour was analysed after conducting the MPTT (see 4.7.3 for discussion of M206's MPTT). During M206's single-housing, F366 was also concurrently isolated for the same purpose, however she did not demonstrate the same change in welfare. This feature highlights what will be shown in the following sections and chapters: the differing nature of welfare at the individual level.

Table 4.8 shows the details of the dogs which were included or excluded from the study. The time from last use in the development of the welfare assessment framework was as follows for each group: SP, eight months; DMPK, five months; Stock, four months.

Table 4.8: Status of dogs in each Group at time of MPTT

Status in MPTT	**SP**	**DMPK**	Stock
Included in MPTT	Ernie, George, Peewee, Ringo	F37, F38, M2, M8, M12, M13	F268, F273, F366, M292, M1
Euthanised	Bert, Ernie, Nibbler, Bouncer	F25, F26	
Excluded			M206

4.3.2 Materials and testing area

4.3.2.1 Testing area

Testing took place in the procedure pod closest to the dog's home pen. While procedure pods were used for both health checks and regulated procedures, dogs

showed clear anticipatory behaviour (low posture, lip licking), suggesting an association between being placed on the table and an aversive event. A quiet area for testing was needed so the procedure pods were the only available location, however it was necessary to ensure that dogs were not 'primed' for an aversive event, as this may have biased recorded thresholds. Dogs appeared to be comfortable when allowed to explore the floor area of procedure pods.

4.3.2.2 The 'Prod'

Research conducted by the manufacturers has suggested that with a tip size >6mm it is almost impossible to manually exert sufficient pressure to cause nociception using the Prod, ensuring that the nociceptive threshold is not approached. The device also provides controlled increases in pressure of 2N sec^{-1} ensuring that a rapid increase in pressure leading to nociception cannot occur. A maximum pressure of 28N and the use of a tip of at least 8mm diameter was agreed with the Home Office inspector as not constituting a regulated procedure. The tip of the 'Prod' was placed centrally on the back of the dog (a less sensitive area than others such as paws or tail), with pressure increasing at a steady rate of 2N sec^{-1} using a 'traffic light' system which tells the user to increase or decrease the rate of pressure. When the dog moves away from the pressure, the 'Prod' records the maximum force applied in Newtons (N).

4.3.3 Testing protocol

Veterinary staff who tested the protocol during ethical approval reported that when restrained, dogs consistently tolerated forces up to the maximum of 28N. It was observed during testing that this was true when dogs were restrained in any manner, as they ceased to respond to the application of the device. This is consistent with the history of the dogs, as conditioning for a lack of response to aversive stimuli (habituation) is considered to be important during dosing, sample collection and physical examination, whether this constitutes desensitisation or a freeze response.

In order to obtain a true measure of MPT it was necessary to allow the dogs to respond to the device which required that dogs were completely unrestrained. This however was problematic as it was difficult to obtain a reading from a dog which moved around during application of the device. A differentiation had to be made between a dog which was moving around because of excitement, because it was attempting to explore the environment or engage with the experimenter, or because it was truly responding to the sensation of the device's application. The methodology which proved to be successful required that the dog was allowed to fully explore the

environment and engage with the experimenter (who did not respond to attempts to elicit attention or play) until it was content to sit or stand still. The dog was also allowed to explore the device, which was presented and gently touched to the dog, demonstrating that it would not cause pain. Testing only began when the dog was content to sit or stand quietly and did not appear to want to interact with the device. The time required for this varied between dogs but typically took at least five minutes.

During the extensive interaction with the dogs for other studies, they had also become accustomed to reading cues from the experimenter, such as raising of an open hand to signal that they should become calm. This signal was presented to the dog before each trial, but did not appear to function in the same way as restraint as they did not cease to respond to the device's application. Adherence to this protocol was critical to obtaining a true reading of MPT, in particular to ensuring readings were not influenced by the dog's response to factors in the environment and the presence of, or handling by, the experimenter.

Consistent with the methodology of several other papers utilising the 'Prod' (e.g. Hunt et al., 2010; Polson, Taylor & Yates, 2012; Bortolami et al., 2013) three readings were taken on each sampling day. As Hoffmann et al. (2010) found that 'mechanical threshold' was stable over six months when testing using similar equipment, testing on three separate days was considered sufficient to gain a valid mean value for MPT. The protocol for testing described above was repeated on each testing day.

During each trial, the device was applied centrally to the dog's back using the 'traffic light' signal (green light to increase force, red light to decrease force) to ensure a consistent rate of application of 2N sec^{-1}. When the dog moved away from the direction of force, the device displayed the maximum force applied in N to two decimal places. This reading was recorded and the protocol repeated.

4.3.4 Ethical approval

In addition to approval by University and Industrial Partner ethical committees, the use of the 'Prod' was reviewed by the local Home Office committee in conjunction with veterinary staff. Other apparatus for testing nociceptive threshold had been considered to constitute a regulated procedure under A(SP)A due to the application of a stimulus designed to induce pain. Upon discussion of the methodology of the study, along with the information on nociceptive thresholds provided by TopCat Metrology Ltd, it was agreed that the 'Prod' could be used as described with a maximum force of 28N to be applied with the use of the 8mm tip. Using the information from TopCat Metrology Ltd and having tested the protocol on several members of scientific staff and then on

dogs which were not participating in this study, it was decided that up to a maximum of 28N, what was being tested was indeed sensitivity to pressure rather than nociception. Dogs were also considered by veterinary staff to be healthy and not suffering from any condition which would cause increased sensitivity to a mechanical stimulus.

4.4 Analysis

4.4.1 Cognitive bias testing

Data were tested for normality and were found to be normally distributed within-dog. Data were then analysed using repeated-measures ANOVAs and planned post-hoc t-tests to examine the differences in time to run to the stimulus boxes. This was to determine if the results support the use of the training criterion as an adequate measure of learning and attempt to identify patterns suggesting an 'optimistic' or 'pessimistic' bias.

4.4.2 MPTT

There were nine readings available for each dog (three from each testing day). Data were not normally distributed. Between-group analysis was conducted using a Kruskal-Wallis independent samples test, and between-affective state analysis was done using a Mann-Whitney U test.

4.5 Results of cognitive bias training and testing

4.5.1 Training

The difference in latency to rewarded (R) and unrewarded (UR) boxes in testing was of particular interest as it would be expected that dogs which reached criterion in training would consistently run faster to the R box than the UR box. If dogs failed to maintain this difference in testing, the results of testing could not be considered valid. A paired-samples t-test was performed and the result was found to be highly significant for seven dogs and significant for all other dogs. Tables 4.9 - 4.11 shows the results of these t-tests.

Table 4.9: T-tests on latencies to R and UR boxes in training for SP dogs

Dog	**R mean (s)**	**UR mean (s)**	t(35)	**p**
Bert	1.87	3.02	6.62	<.001
George	1.58	2.17	10.27	<.001
Bob	2.55	13.77	5.62	<.001
Peewee	2.28	4.51	2.53	<.001
Ringo	1.75	22.43	10.27	<.001

Table 4.10: T-tests on latencies to R and UR boxes in training for DMPK dogs

Dog	**R mean (s)**	**UR mean (s)**	t(35)	**p**
F25	1.47	3.06	2.09	.044
F26	1.11	11.61	4.74	<.001
F37	10.82	18.70	2.36	.024
M12	3.01	7.25	2.25	.031
M13	3.31	8.79	2.66	.012

Table 4.11: T-tests on latencies to R and UR boxes in training for Stock dogs

Dog	**R mean (s)**	**UR mean (s)**	t(35)	**p**
F268	1.56	5.16	2.83	.008
F273	1.23	3.56	2.09	.044
F366	1.29	6.91	3.15	.003
M206	1.61	8.04	3.54	.001
M292	2.26	10.98	3.92	<.001
M1	1.40	3.24	2.285	.029

4.5.2 Testing

A repeated measures ANOVA was conducted to determine if there were significant differences between times to the three probe boxes. Where the latency to probe boxes was not significantly different, this suggests that dogs responded to probe boxes in a consistent manner, regardless of whether it was closer in shade to R or UR boxes. There were only significant differences in latencies for one dog, Ringo ($F(2,10)=8.144$, $p=.008$).

Across the three groups (SP, DMPK and Stock), there were two common patterns of response: an 'optimistic' response in which dogs consistently ran quickly to all but the unrewarded box, regardless of whether or not it had previously been rewarded, and a 'pessimistic' response in which dogs approached either only the rewarded box, or approached the other boxes at much longer latencies. Figure 4.4 shows the pattern of responses typical of an 'optimistic' dog while Figure 4.5 shows the pattern of responses typical of a 'pessimistic' dog. The dogs which were unable to overcome their neophobia

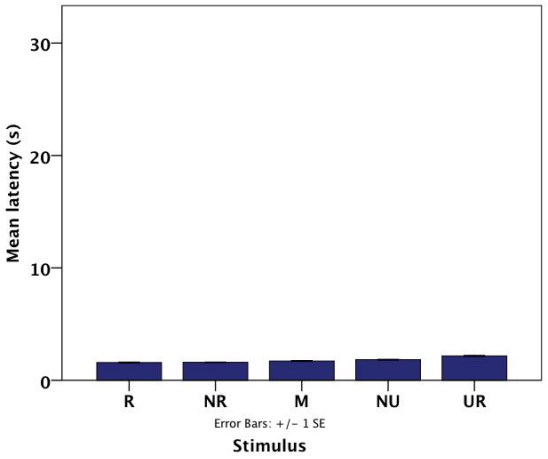


FIGURE 4.4: Mean latency to each stimulus box for George

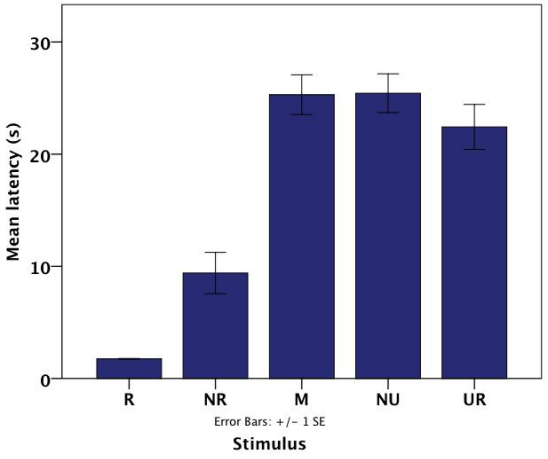


FIGURE 4.5: Mean latency to each stimulus box for Ringo

to progress to testing were classified as 'pessimistic'; they did not approach the 'novel' stimuli of the rewarded and unrewarded boxes and the literature suggests that this represents the use of the 'withdraw' mechanism of the 'approach/withdraw' paradigm and a tendency to withdraw from unfamiliar, or ambiguous, stimuli, which is what the cognitive bias paradigm is testing. The results of individual testing can be found in Appendix A.

Figure 4.6 shows the prevalence of 'optimistic' and 'pessimistic' types across the three groups; there was an equal distribution of optimists and pessimists in SP and DMPK groups, while only one Stock dog showed a pessimistic style. This supports the hypothesis that increasing time in the unit and experience of regulated procedures have a negative impact on affective state.

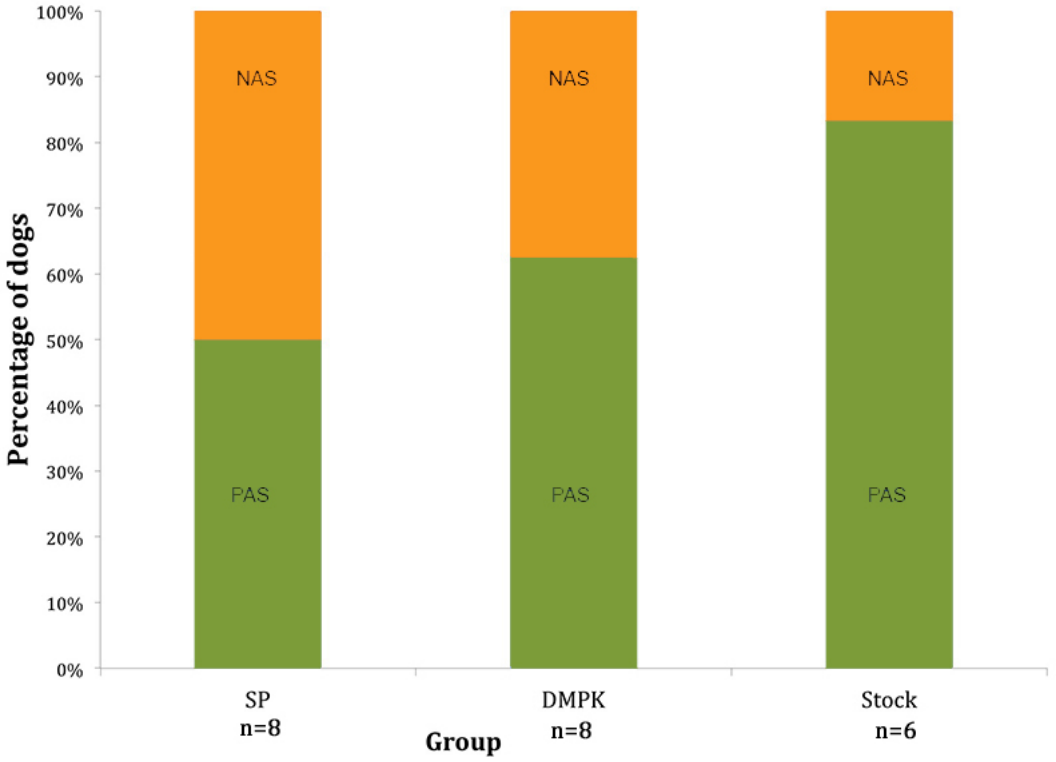


FIGURE 4.6: Distribution of positive and negative cognitive biases across groups

4.6 Results of MPTT

There were three comparisons of interest: between-group, between-affective states and between-sex. To test between-group differences, a Kruskal-Wallis independent samples test was utilised, and for between-affective states and between-sex differences a Mann-Whitney U test was utilised. MPT is expressed in Newtons (N).

4.6.1 Between-Groups differences

The results of MPTT for each group is shown in Table 4.12.

TABLE 4.12: Results of MPTT by Group

Group	n	Mean (N)	SD
SP	4	16.78	5.94
DMPK	6	16.76	3.86
Stock	5	20.89	5.37

The Kruskal-Wallis independent samples test showed that there was no significant difference in MPT between groups, χ^2=1.583, p=.453, although Stock dogs had a higher mean MPT compared to SP and DMPK groups,

4.6.2 Between-affective states differences

Table 4.13: Results of MPTT by Affective State

Affective State	n	Mean (N)	SD
PAS	9	21.58	2.64
NAS	6	12.99	2.50

The Mann-Whitney U test showed a significant difference in MPT between Affective States, U=0.00, Z=3.182, p<.001.

4.6.3 Between-sex differences

Table 4.14: Results of MPTT by Sex

Sex	n	Mean (N)	SD
Male	10	17.64	4.84
Female	5	19.14	5.79

The Mann-Whitney U test showed no significant difference in MPT between sexes, U=23.00, Z=.245, p=.806.

4.7 General Discussion

4.7.1 Training for cognitive bias testing

4.7.1.1 SP dogs

Training for this study proved to be problematic. All dogs showed what was initially a surprising level of neophobia towards the novel environment, novel experimenter and in particular, the novel stimulus boxes. When considering the level of consistency in the dogs' daily routine, this may be less surprising. Husbandry and procedures were conducted by a small number of familiar staff and the dogs were rarely taken outside of their immediate environment. Further, unfamiliar equipment may have been associated

with procedural work. The experimental procedure was completely unfamiliar to these dogs prior to testing and so a greater number of familiarisation sessions should be considered for conducting similar studies with dogs which do not regularly encounter unfamiliar people. The environment did contain pens similar to those in which the dogs are normally housed and although noises sometimes disturbed training and testing, the dogs did not seem unduly upset by this. No positive reinforcement training was routinely administered to SP dogs and so the dogs were not used to taking food as a reward, the length of time to acclimatise to this was also underestimated.

Two of the eight dogs exhibited such a high level of neophobia that training was terminated and another during testing. Whilst it not possible to draw conclusions about the affective state of these dogs from testing, it is clear from their responses to the unfamiliar training situation that they viewed unfamiliar or ambiguous stimuli and events as potentially threatening, which would suggest a negative outlook.

4.7.1.2 DMPK dogs

DMPK dogs proved much easier to train than SP dogs and showed considerably fewer signs of neophobia. This may be in part due to modifications to the study design. Following the problems encountered with SP dogs, the data collection for Chapters 5 and 6 was collected before CB took place to encourage dogs to become more familiar with the experimenter. Extra steps were taken to ensure that dogs were comfortable with the experimenter before being removed from their housing zone. However, given the differences in husbandry practices for DMPK dogs, the more frequent staff contact and activity in the housing zone is likely to have resulted in more resilient dogs. The dogs were required to travel a distance to the testing area and the dogs which proceeded from familiarisation to training were able to do this without problems. Using transport boxes elicited a negative response from the dogs, likely due to past associations, so allowing the dogs to walk on the lead also increased the amount of control they had over participation. SP dogs exhibited signs of anxiety when removed from their housing zone and it is unlikely that any amount of training would have lead to them being comfortable walking in an open area on a lead.

Despite these differences, there were still three dogs who were unable to take part due to high levels of neophobia. While the overall level of neophobia appears to be lower in this group, the same number of dogs were unable to participate.

4.7.1.3 Stock dogs

When comparing this study to Burman et al.'s (2011) study from which the design was adapted, it is apparent that comparatively fewer trials were needed to reach criterion in this study. Burman and colleagues found that only two of 12 dogs took less than 40 trials, with five requiring over 100. Of the total of 16 dogs which proceeded to testing, 11 completed training in under 40 trials. Three dogs required fewer than 20 trials.

Whilst Burman et al. (2011) describe the familiarisation period for a maze also used in the study, there is no mention of familiarisation for the equipment and area used for cognitive bias testing, perhaps because it was not deemed necessary for their sample. The need for familiarisation in this study may account for the difference. The different populations from which dogs were drawn may also have influenced their learning.

Laboratory-housed dogs have little interaction and stimulation in their home environment compared to pet dogs or shelter dogs which often receive regular exercise or training with staff. These dogs, at least those which were not prevented from participating by their own neophobia, quickly became focused on the task and appeared to find participation in the task rewarding in its own right. This feature of the dogs' behaviour was noticeable in other aspects, particularly in the dogs which were lead walked for transport to the testing area and those which quickly picked up on cues. As opportunities to interact with staff were limited it appears that the dogs were very focused on any opportunities presented to them which allowed them to interact with a person or task. This is a natural behaviour in domesticated dogs and is likely to be present more so in dogs selectively bred for temperament and cooperation with humans. Laboratory animals also experience few opportunities to exercise choice in their environment, dogs were never forced to participate in any aspect of the study and all aspects from removal to the testing area to the task to return to the home pen were to an extent within the dogs' control. Choice is highly valuable to laboratory animals and this is likely to have increased their willingness to participate and engage.

This study also used a distance of 3m from holding area to box, half that used in Burman et al.'s (2011) study. In piloting, it was found that dogs did not get close enough to the box at a distance of 6m to discover the food reward. While the use of a smaller distance did not affect results (the same criterion of at least 0.5 seconds difference in speeds to R and UR boxes was used in training) this may have increased the speed of learning. Burman et al.'s (2011) dogs may have required more training because of a lower frequency of contacts with the stimulus box meaning that learning took longer

4.7.2 Cogntive bias testing

For the findings of the testing to be valid, it is first necessary to ensure that the methodology was capable of producing valid results. There are two findings which suggest this. First, all dogs demonstrated a significant difference in latencies between R and UR boxes. Had the criterion used in testing not been sufficient to determine the point at which dogs could determine which box was rewarded and which was unrewarded, this result would not have reached significance. The second is the presence of significant differences in latencies to the probe boxes for all but one dog. Had the dogs not been able to visually discriminate between the stimulus cards, they would not have been able to associate a stimulus card with reward and so this result would not have reached significance.

There are possible variations on the design of the task used in this study. One possibility is to use a mildly aversive food or other stimulus in the unrewarded box, however in this case as dogs were initially nervous and hesitant in response to the equipment and reticent to take food, this may have extinguished any searching response and prevented training proceeding. The other, as recommended by Mendl et al. (2009), is to require an active response to both stimuli to prevent the confounding effect of lack of motivation. In this case, there were a limited number of approved foods available for use as rewards and as the dogs failed to respond at all to the dry diet, there was only one food stuff suitable for use.

Whilst there was some variation in the differences in latencies to the boxes between dogs in testing, there seems to be a clear pattern of positive bias across all but two dogs. Several of the dogs showed a consistent fast running speed to all boxes (around 2.5s), even those which had never been rewarded. Given that the dogs had reached criterion in training and showed a significant difference in latencies to R and UR boxes in testing this can only be because of an ‘optimistic belief that the boxes may at some point be rewarded. Due to the terms ‘optimistic and ‘pessimistic having somewhat confusing colloquial meanings, dogs with an ‘optimistic pattern of responses were categorised as positive affective state (PAS), while dogs with a ‘pessimistic’ pattern of responses were categorised as negative affective state (NAS).

4.7.3 MPTT

One feature of mechanical pressure threshold training which must be discussed is the methodology needed to obtain true readings for all dogs. M206 was the only dog available for MPTT who was excluded from the sample on the basis of changed welfare

since cognitive bias testing was conducted. Although his MPT could not be included in analysis, the readings are still of interest. In comparison to the other dogs, who moved away from the Prod when the sensation elicited became uncomfortable, M206 failed to respond and almost every trial reached the maximum force of 28N before abandonment. In contrast to the NAS dogs that showed consistently lower MPTs than PAS dogs, M206's decreasing welfare seems to have brought about learned helplessness and a 'freezing' response. This is supported by his behaviour during testing, with lowed body posture, ears and tail and a lack of responsiveness to the environment similar to that observed in dogs undergoing aversive procedures. This apparent contradiction with the NAS dogs may be due to the proximate timing of the aversive event which appears to have decreased his welfare. M206 was tested within days of being group-housed following three weeks of single-housing. The negative affective state observed in some dogs may be a stable characteristic over time as it does not appear to have changed. M206's decreased welfare may be an acute response to a stressor which may or may have been transient, it is impossible to know this without having observed him over time following his release from single-housing.

One final feature of MPTT is its support of cognitive bias testing's measurement of affective state. As a lower nociceptive threshold is known to predict a negative affective state, it was predicted that if cognitive bias testing had truly measured affective state that those individuals with a PAS as determined by cognitive bias would have a lower MPT than NAS individuals. This proved to be true, with PAS dogs having a significantly higher threshold than NAS dogs, suggesting that cognitive bias 'optimism' and a positive affective state (PAS) should be considered to be closely linked, if not the same phenomenon.

4.7.4 Conclusions - affective state

Cognitive bias testing was developed as a means of understanding and quantifying the emotional state of non-verbal animals. While some debate exists around the underlying mechanism (Mendl et al., 2011), it is clear from the findings of this study that some factor separated PAS and NAS dogs which manifested itself in response to cognitive bias testing and mechanical pressure threshold testing. PAS dogs were characterised by lower arousal in response to novel environments, equipment and the experimenter, as well as in response to the test of mechanical pressure threshold. NAS dogs showed an opposite pattern. While neither CB or MPTT can ever give us a true insight into the unseen emotional states of animals, it seems unlikely in view of what is currently known about emotions and their role in responding to threatening stimuli that dogs exhibiting the patterns seen in NAS dogs could nonetheless have positive

emotional states. Indeed, as has been clearly stated by authors such as (Dawkins, 2008b), it is not possible to have good welfare while having a negative emotional state. A subset of dogs with a different pattern of welfare has clearly been identified using the tests in this study. It should also be clear, when comparing the histories of the dogs (Chapter 3) that factors relating to regulated procedures, housing practices or staff contact may have influenced these differing patterns of welfare. However neither CB or MPT testing are practical to carry out in a welfare assessment in a busy environment such as an animal unit, neither do they allow any insight into the effects of welfare on the quality of data output obtained from the dogs. The next step in developing the Welfare Assessment Framework therefore deals with these issues.

CHAPTER 5

Behaviour and cardiovascular function of dogs in the home pen

Developing the Welfare Assessment Framework II

"The scientific study of animal welfare should be promoted so that decisions are made on factual rather than emotional grounds"

Donald Broom (1991), pg. 5174

Abstract

In Chapter 4, the affective states of the dogs were measured. Two distinct groups were formed on the basis of testing: Positive Affective State (PAS) and Negative Affective State (NAS). Affective states were found to be differently represented between the three groups of dogs: SP, DMPK and Stock. In this chapter, a 'baseline' of behaviour and cardiovascular parameters is developed from observations in the home pen. The aim is to determine which behaviours are most suitable to include in a welfare assessment framework and are most sensitive to different welfare states. The effects of group, sex and affective state on these are investigated to determine which factors might influence welfare.

5.1 Introduction

Measuring affect in Chapter 4 provided a measure of the internal state of the dogs, which is key to understanding welfare. However the length of time and the complexity involved in conducting cognitive bias testing means it is neither practical nor feasible in the laboratory environment. One of the primary aims of this project is to develop a tool for use by technicians which provides the same measurement of welfare as that gathered in the development of the Welfare Assessment Framework, while being easy-to-use for technicians and care staff. Frameworks have previously been formed for nonhuman primates (Wolfensohn & Honess, 2008), in particular macaques (Tasker, 2012).

While positive affect and positive welfare can be considered to be co-existing (Dawkins, 1990), this is not sufficient to provide a basis for investigating quality of data output. It is necessary to measure data output directly. Some of the characteristics of data quality which are of most interest to us are related to precision, repeatability, variation and accuracy of measurement (Chapter 2, Appendix I). In order to make precise and accurate measurements, there is a requirement for confounding factors to be controlled (see Appendix I). Measures affected by increased variance which cannot be measured accurately will be subject to poor repeatability. In safety assessment testing, physiological measures such as body weight and food consumption will be part of the standard battery of measures (Gad, 2006). In addition, one of the primary uses of the dog model is for cardiovascular function, part of the safety pharmacology battery of measures being to detect (often subtle) effects on heart rate and blood pressure (ICH, 2008).

Because the terms of the project grant meant that only non-invasive and non-regulated procedures were to be used to gather information on welfare and quality of scientific output, blood sampling was discounted. Cortisol is an ambiguous indicator of welfare and so salivary cortisol collection was also discounted. Food consumption and body weight were regularly collected by the responsible carer and continuous telemetered cardiovascular data could be obtained remotely, therefore these were chosen as appropriate measures of data quality.

Since these data (behavioural, affective, cardiovascular, physiological) had not been examined this integrative way previously in the literature, data collection and analysis here is somewhat exploratory, identifying patterns of behaviour, physiology and cardiovascular parameters at 'baseline' in the home pen. These data will then examined for differences between groups, affective states and sexes to determine how these factors might relate to welfare.

5.2 Methodology

The methods of data collection for behavioural and cardiovascular data are described in Chapter 3. For each of the three groups of dogs, behavioural data were recorded in 10 five-minutes observation sessions in the home pen, while SP group had simultaneous recording of CV data. A full description of the behaviours recorded can be found in Appendix B.

5.2.1 Physical data

For all dogs, measures of physical health were obtained in the form of regular body weight and food consumption data. Body weights were obtained weekly by care staff during health checks. The weight in kilograms was recorded on each dog's clinical observation sheet. In addition, SP dogs only had daily food consumption recorded, however DMPK and Stock dogs were group-fed and daily food consumption was not recorded. DMPK dogs were singly-housed for feeding while on-study, and food consumption recorded. These data collected during studies were not a reliable representation of baseline food consumption and so were not used in this study. It would not have been appropriate to restrict the dogs to individual housing for the purpose of measuring food consumption.

5.3 Results of behavioural analysis

5.3.1 Summary of results

Tables 5.1-5.3 highlight the significant findings of between-subject baseline analysis of behaviour and cardiovascular function. There are three predominant sets of findings: between-groups (5.1), between-affective states (5.2) and between-sex (5.3). There are also interactions between these factors which are depicted in these tables.

Table 5.3 shows differences found between sexes. Tables 5.2 and 5.3 also show interactions between other factors (i.e. affective state, group or sex). For all tables, upwards (↑) arrows indicate a group showing higher levels of a behaviour, downwards (↓) arrows indicate a group showing lower levels of a behaviour and '**X**' indicates groups which did not show an increased or decreased level of a behaviour. Colour coding shows whether the change in the indicated behaviour demonstrates increasing or decreasing welfare (decreasing welfare; increasing welfare).

TABLE 5.1: Significant findings of between-Groups analysis of baseline behaviour and CV

Behaviour Type	Behaviour	SP (n=8)	DMPK (n=8)	Stock (n=6)
Location	Front	X	X	↑
	Rear	↑	X	X
Positive Welfare Indicators	Resting head down	X	↓	X
Negative Welfare Indicators	Stand against walls	↓	X	↑
Posture	High posture	X	↓	X
	Neutral posture	X	↓	X

TABLE 5.2: Significant findings of between-AS analysis of baseline behaviour and CV

Effect or interaction	Behaviour type	Behaviour	PAS (n=13)	NAS (n=9)
Affective state	Location	Barrier	↓	↑
	Positive Welfare Indicators	Resting head up	↑	↓
	Negative Welfare Indicators	Stand alert	↓	↑
	Posture	High posture	↓	↑
	CV	SBP (n=8)	↓	↑
		DBP (n=8)	↓	↑
Group*Affective state			**SP DMPK Stock**	**SP DMPK Stock**
		High posture	**X** ↑ **X**	**X** ↓ **X**

TABLE 5.3: Significant findings of between-sex analysis of baseline behaviour and CV

Effect or interaction	Behaviour type	Behaviour	Male (n=7)	Female (n=7)
Sex				
	Positive welfare indicators	Resting head down	↓	↑
	Negative welfare indicators	Stand alert	↑	↓
	Posture	High posture		
		Tail wagging (high)	↓	↑
Group*Sex			**DMPK Stock**	**DMPK Stock**
	Positive welfare indicators	Amicable	↑ ↓	↓ ↑
		Interact with environment	↑ ↓	↓ ↑
	Negative welfare indicators	Stand against walls	↓ ↑	↑ ↓
		Sit alert	↓ ↑	↑ ↓
	Posture	High posture	↓ ↑	↑ ↓
		Neutral posture	↑ ↓	↓ ↑
		Tail wagging (high)	↓ ↑	↑ ↓
	Events	All events	↓ ↑	↑ ↓
Affective State*Sex			**PAS NAS**	**PAS NAS**
	Negative welfare indicators	Sit alert	↑ ↓	↓ ↑
	Posture	Neutral posture	↓ ↑	↑ ↓
		Tail wagging	↑ ↓	↓ ↑

5.3.2 Behavioural analysis of all dogs (n=22)

The first step in the analysis of baseline behaviour was to determine which behaviours occurred for more than 5% of time for at least 50% of dogs. These criteria were used to identify behaviours which occurred at a sufficiently high level to be easily observable and suitable for inclusion in the welfare assessment framework. The table shows the means and standard deviations of behaviours meeting these criteria, a total of 18 locations, states and events.

Table 5.4: Home pen behaviours selected for analysis by type, showing means and standard deviations

Behaviour type	**Behaviour**	**Mean % of time**	**SD**
Location	Front of pen	42.50	26.77
	Back of pen	29.44	29.62
	Barrier	8.23	11.12
Positive welfare indicators (PWIs)	Amicable behaviours	5.98	8.76
	Interacting with environment	7.09	9.46
	Resting head down	15.20	23.46
	Resting head up	16.87	14.56
	Calm locomotion	4.60	7.60
	All play behaviours	5.57	15.04
Negative welfare indicators (NWIs)	Standings against walls	6.23	9.80
	Sitting alert	23.29	18.21
	Standing alert	14.93	14.94
Postures	High posture	25.92	18.86
	Neutral posture	46.85	22.39
	Half-low posture	10.34	13.50
	Low posture	13.34	17.84
	Tail wagging high	15.77	15.92
Events	All events	5.76	14.14

5.3.2.1 Location within the home pen

There was a significant difference in time spent in each of the three locations ($F(2, 148)=56.59$, $p<.001$), with dogs preferring to be located at the front of the pen. Dogs spent significantly more time at the front than at the rear ($t(150)=3.011$, $p=.003$) and more time at the rear than at the barrier ($t(150)=14.11$, $p=<.001$). Location in the home pen in combination with the behaviour expressed is likely to reflect welfare as the dog's proximity to and ability to observe other dogs or events changes between locations.

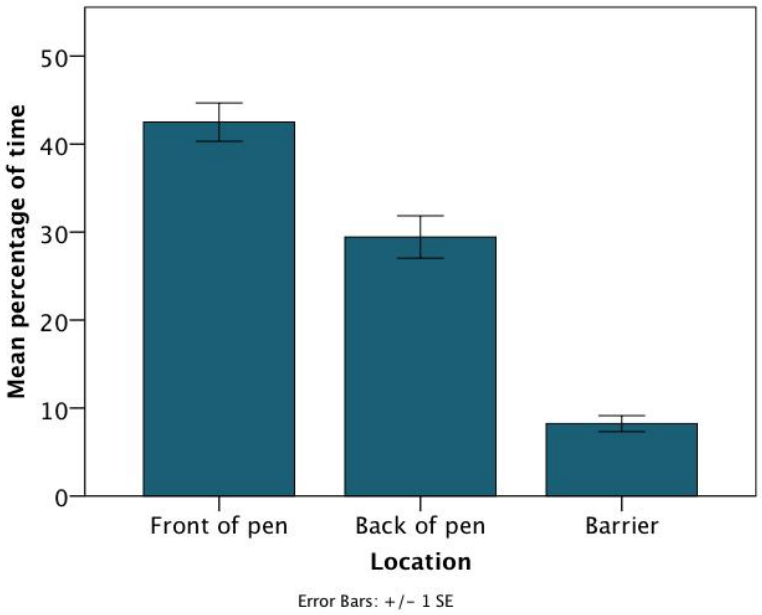


FIGURE 5.1: Mean percentage of time spent in each location within the home pen

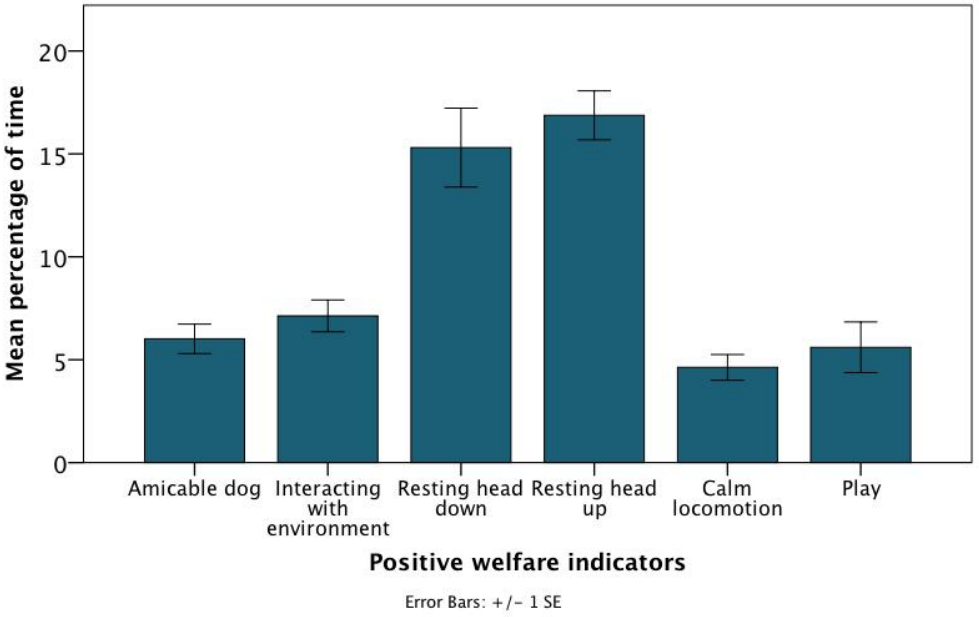


FIGURE 5.2: Mean percentage of time spent in positive welfare indicating behaviours in the home pen

5.3.2.2 Positive and negative welfare indicators

Two of the most prevalent behavioural states were the resting and alert behaviours, together accounting for 70.3% of time. These positive welfare-indicating (PWI) and negative welfare-indicating (NWI) mutually exclusive states account for over 95% of time, meaning that less than 5% of behaviours were excluded due to not meeting

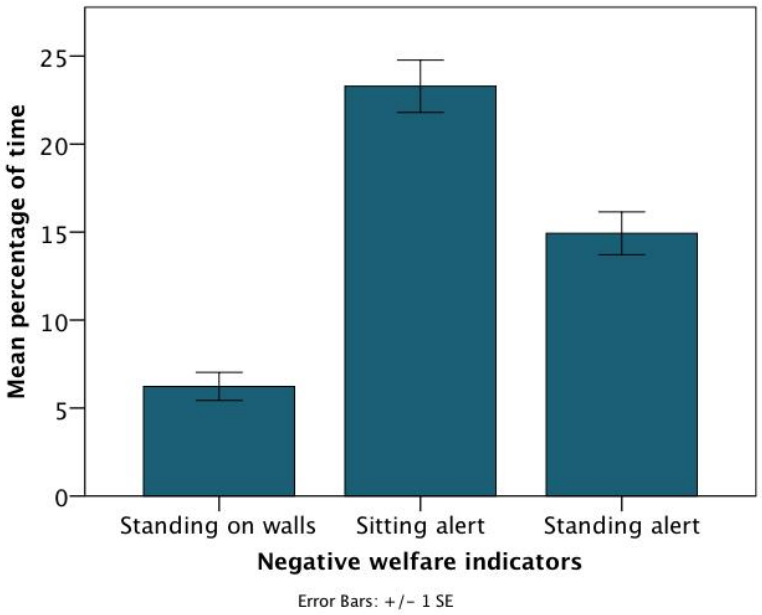


FIGURE 5.3: Mean percentage of time spent in negative welfare indicating behaviours in the home pen

criteria. This also suggests that resting and alert behaviours may be the best indicators of positive and negative welfare because they occur most frequently; therefore the balance of behaviour between alert and resting behaviours may be a useful indicator of welfare.

When comparing the total time spent in positive (n=6) and negative (n=3) welfare indicating behaviours shown in Table 5.4, as a whole dogs were spending more time in PWIs than NWIs ($t(149)=2.495$, $p=.014$). There is also a greater diversity of PWIs when only looking at these behaviours. A greater variety in behaviour (number of behaviours displayed) is considered to indicate greater welfare, particularly a greater variety of PWIs and a lower variety of NWIs.

5.3.2.3 Postures

Posture provides information about dogs' responses to stimuli in the environment. High posture can indicate excitement, vigilance or aggression, while low posture can indicate fear, anticipation or attempts to placate when directed at humans or other dogs. A neutral posture suggests that dogs are neither anticipating or reacting to something in the environment. While high and low posture may been seen in direct response to a stimulus, for example on the presentation of food or in anticipation of an aversive event, seeing these posture exhibited throughout the day in the absence of such stimuli may indicate decreasing welfare.

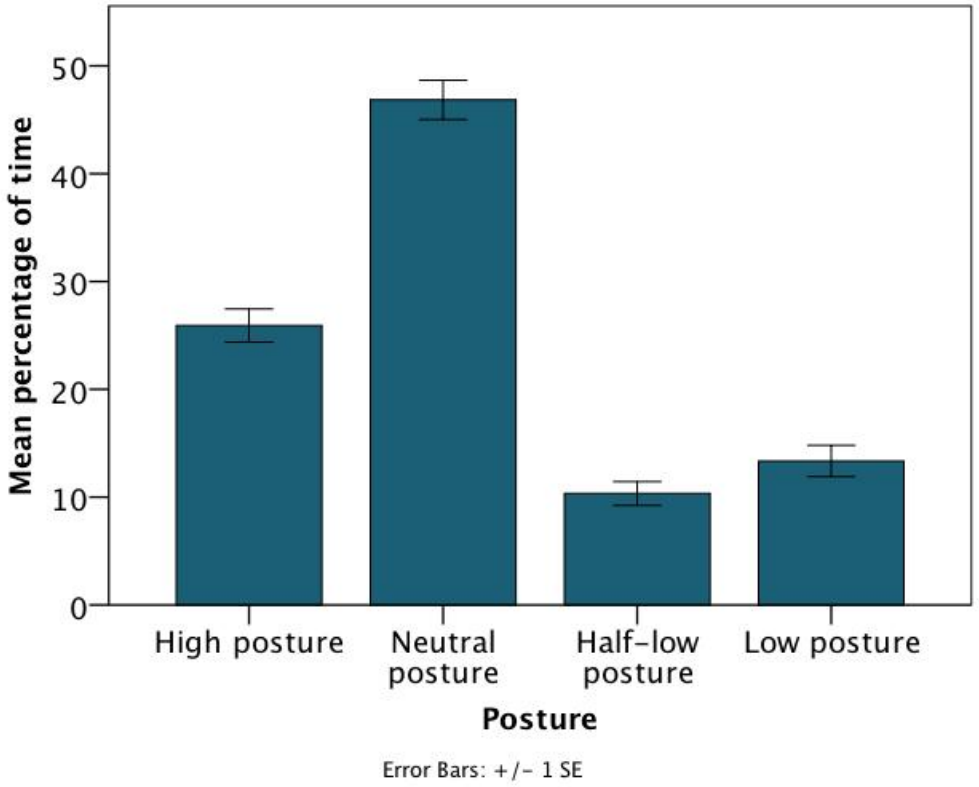


FIGURE 5.4: Mean percentage of time spent in high, neutral, half-low and low postures in the home pen

Across all groups, there was a significant difference in time spent in each of these four postures, with dogs spending nearly 50% of time with neutral posture. Time spent with high posture was significantly less than with neutral posture ($t(150)=7.021$, $p<.001$), and low posture was displayed significantly less than high posture ($t(150)=14.147$, $p<.001$). There was also a significant difference between low and half-low postures ($Z=4.028$, $p<.001$), with low posture occurring infrequently.

5.3.2.4 Behavioural events

All of the behavioural events listed in the coding scheme in Appendix B have the potential to be considered stereotypic behaviours and therefore negative welfare indicators. It was also anticipated that the frequency of occurrence of these behaviours would be low at baseline as they are more likely to occur in response to a stimulus.

The mean rate of all behavioural events was just less than six per hour, which equates to one behavioural event exhibited every 10.4 minutes. The rate of behavioural events displayed here may be affected by an observation bias which resulted in few behavioural events being recorded for SP dogs. The CCTV cameras located in the

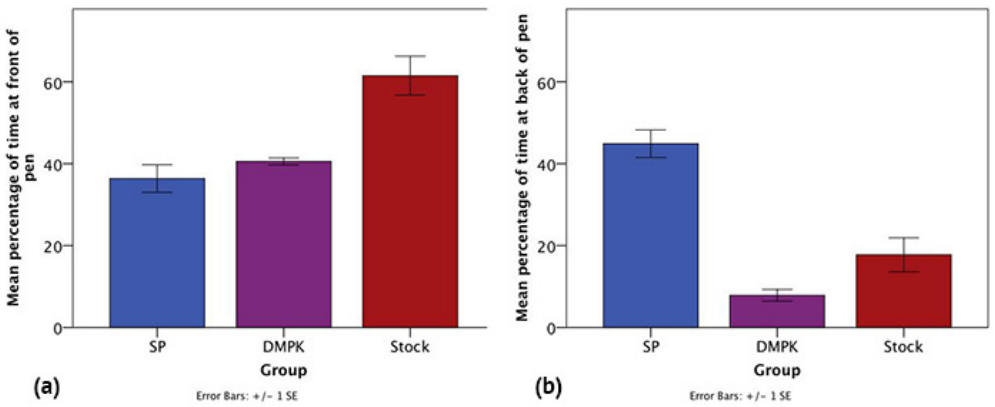


FIGURE 5.5: Mean percentage of time spent at (a) front and (b) back of the home pen

ceiling had a lower resolution and frame rate than the camcorders used later for DMPK and Stock; this together with the location directly above the pen made it difficult to observe short-duration or sudden behaviours, in particular those involving the face such as panting or lip licking. The difference in occurrence of these behaviours in response to Challenges (Chapter 6) in comparison at baseline will elucidate their meaning as welfare indicators.

5.3.3 The effects of Group

Comparing the three groups (SP, DMPK and Stock) was of particular interest because of the differences in a number of factors such as husbandry, housing and history of regulated procedures.

TABLE 5.5: Results of ANOVAs showing significant between-Group differences in home pen behaviour

Effect	Behaviour type	Behaviour	F(2, 148)	p
Group	Location	Front of pen	11.200	<.001
		Back of pen	27.850	<.001
	PWIs	Resting head down	4.450	.013
	NWIs	Standing against walls	18.431	<.001
	Posture	High posture	4.624	.011
		Neutral posture	15.361	<.001

Differences in behaviour, especially PWIs and NWIs, may elucidate the role of these factors in determining welfare, particularly as affective state was found to vary between groups as measured through cognitive bias testing. All behaviours shown in Table 5.4 were examined for between-groups differences. Table 5.5 shows the six

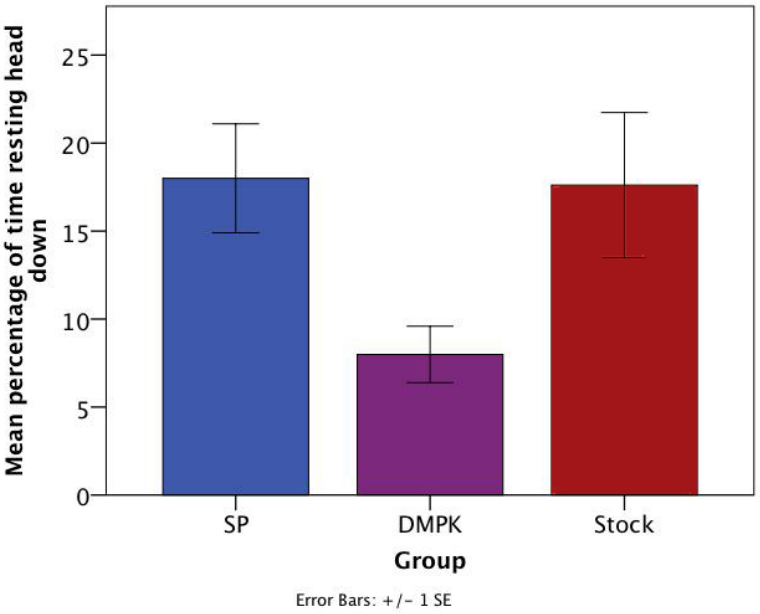


FIGURE 5.6: Mean percentage of time spent resting head down in the home pen

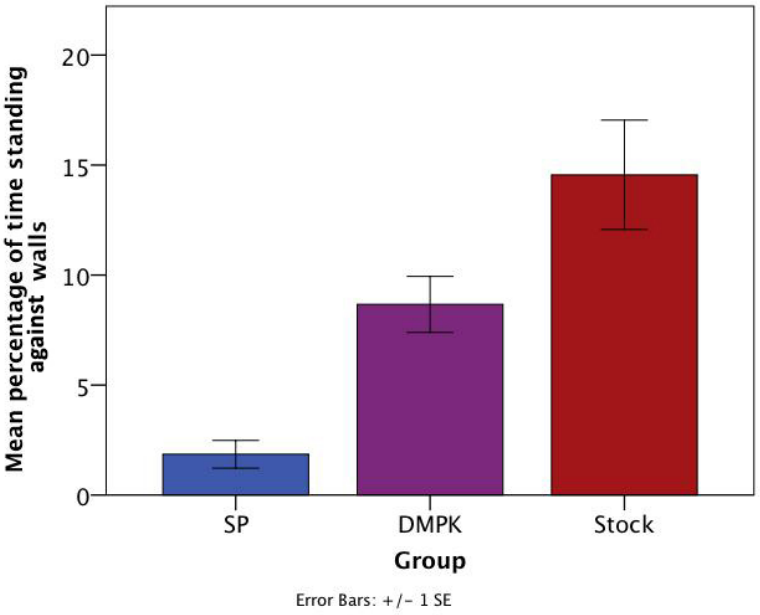


FIGURE 5.7: Mean percentage of time spent standing against walls

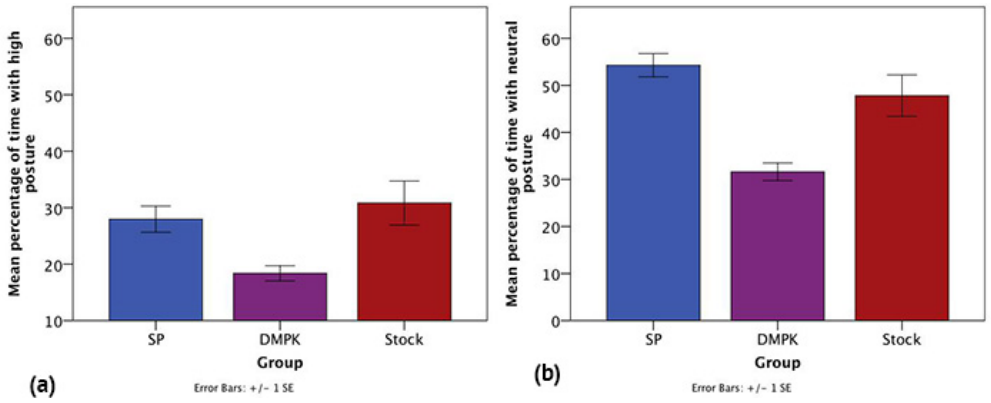


FIGURE 5.8: Mean percentage of time spent with (a) high and (b) neutral posture by Group

behaviours which varied significantly between groups. Interactions between group and other factors (AS, sex) are discussed later in the relevant sections.

SP dogs spent more time at the back of the pen than other groups, while Stock dogs were spending more time at the front than other groups. Although spending time at the back should indicate positive welfare and time at the front should indicate negative welfare, in this case it is likely to indicate the level of activity in the housing zone, with no interruptions in the SP zone and regular activity in Stock. Stock dogs also showed a more positive cognitive bias than SP dogs which suggests that their welfare was not more negative.

DMPK dogs spent the least time resting head down. This does not seem to be reflected in an increase in NWIs such as standing against walls (Fig. 5.7) or having high posture (Fig. 5.8), suggesting that DMPK dogs were active, but not engaged in behaviours indicating high levels of arousal. Stock dogs spent the greatest time standing against walls, another indicator of agitation or excitement, and of negative welfare. As with time spent at the front, in Stock dogs this is likely to be an indicator of the high level of activity in their housing zone and of the higher energy levels of young dogs as other indicators suggest that their welfare was positive. DMPK dogs exhibited both less high and less neutral posture than the other groups. As time with high or neutral posture account for 72% of time, this result is likely to reflect DMPK dogs exhibiting a greater variety of postures although time with half-low or low posture was not significantly different between groups.

5.3.3.1 Number of behaviours displayed

The number of behaviours displayed was compared between groups. A one-way ANOVA showed a significant effect of group on the number of PWIs ($F(2, 148)=4.787$, $p=.021$), the result of SP dogs exhibiting more PWIs than Stock dogs ($t(12)=2.825$, $p=.005$). There was no difference between SP and DMPK or DMPK and Stock. There was also an effect of group on total behavioural states displayed ($F(2, 148)=8.027$, $p=.003$), with SP dogs displaying more behavioural states than either DMPK ($t(14)=3.113$, $p=.008$) or Stock ($t(12)=3.550$, $p=.004$). This suggests that a greater variety of behaviours at baseline, either positive or negative welfare indicators, may be indicative of a negative welfare state. The analysis of affective state which follows will explore this further.

5.3.4 The effects of Affective State

Table 5.6: Results of ANOVAs showing significant between-Affective State differences

Effect or interaction	Behaviour type	Behaviour	F	df	p
Affective state	Location	Barrier	2.499	1, 148	.014
	PWIs	Resting head up	2.059	1, 145	.041
	NWIs	Standing alert	2.625	1, 145	.010
	Posture	High posture	3.192	1, 145	.002
Group x AS	Posture	High posture	2.807	2, 145	.064

There was a clear split in affective states in both SP and DMPK groups, with 50% of dogs showing a PAS and 50% showing a NAS. Stock dogs showed an overall PAS, with one dog exhibiting a marginally NAS. Due to this split in affective states, analysis of behaviours between affective states was conducted to determine if different patterns of behaviour could be identified which would reliably indicate affective state. A positive affective state should be considered consistent with positive welfare and so affective state is analysed as a potentially central measure to the Framework.

Table 5.6 shows the four behaviours that were significantly different between affective states as well as one interaction between group and affective state which was marginally non-significant. Resting and alert behaviours were the most often seen behaviours at baseline and that they differ between affective states suggests that these two behaviours are sensitive measures of welfare.

NAS dogs spent more time at the barrier PAS dogs. Time at the barrier between pens is a reflection of the number of time a dog moves between pens and may be reflective

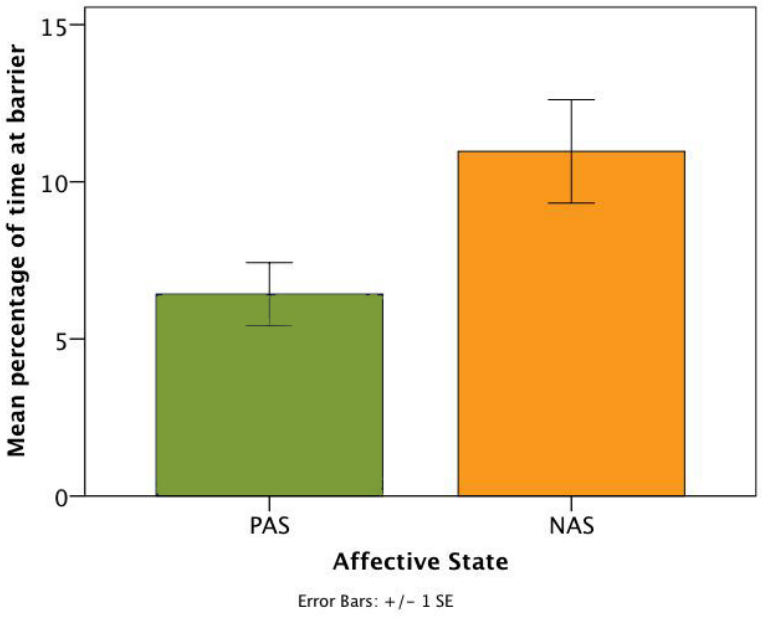


FIGURE 5.9: Mean percentage of time spent at barrier between pens by Affective State

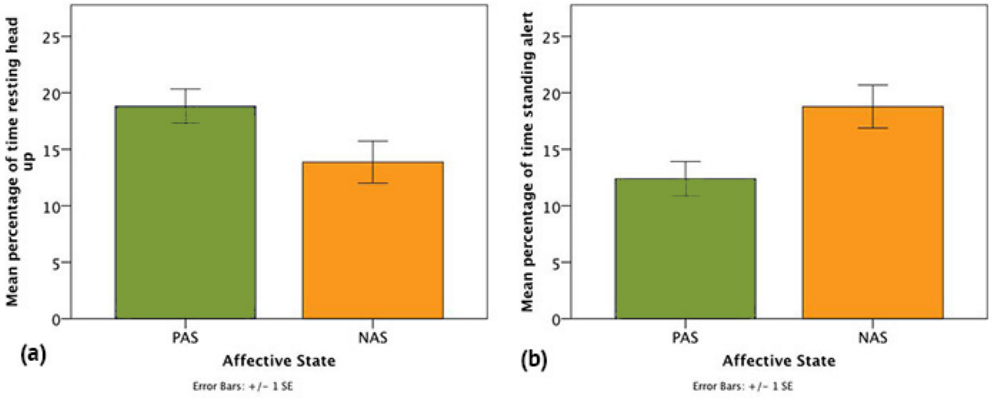


FIGURE 5.10: Mean percentage of time spent (a) resting head up and (b) standing alert by Affective State

of agitated activities such as rapid locomotion or pacing, two behaviours which did not occur for a sufficient duration to be included in analysis. This suggests a higher level of agitation in NAS dogs.

PAS dogs spend significantly more time resting head up than do NAS dogs. Resting head up is a PWI, as it indicates that dogs are awake but resting calmly rather than being alert and orientated to a stimulus. In contrast, NAS dogs spend more time standing alert than PAS dogs, indicating that they are frequently attending to something outwith the home pen rather than resting calmly.

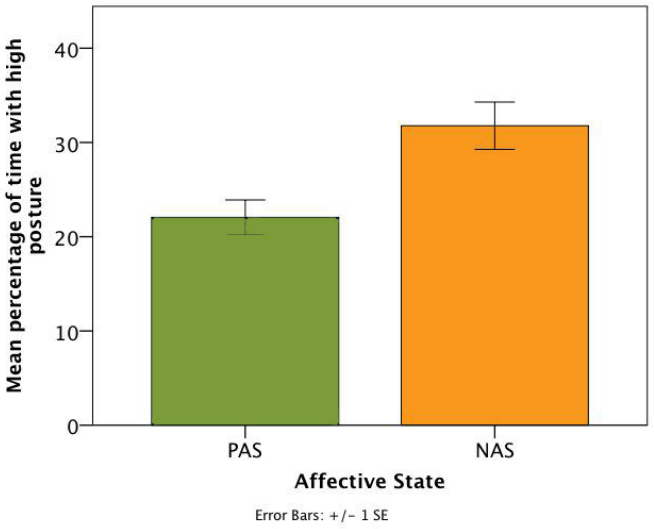


FIGURE 5.11: Mean percentage of time spent with high posture by Affective State

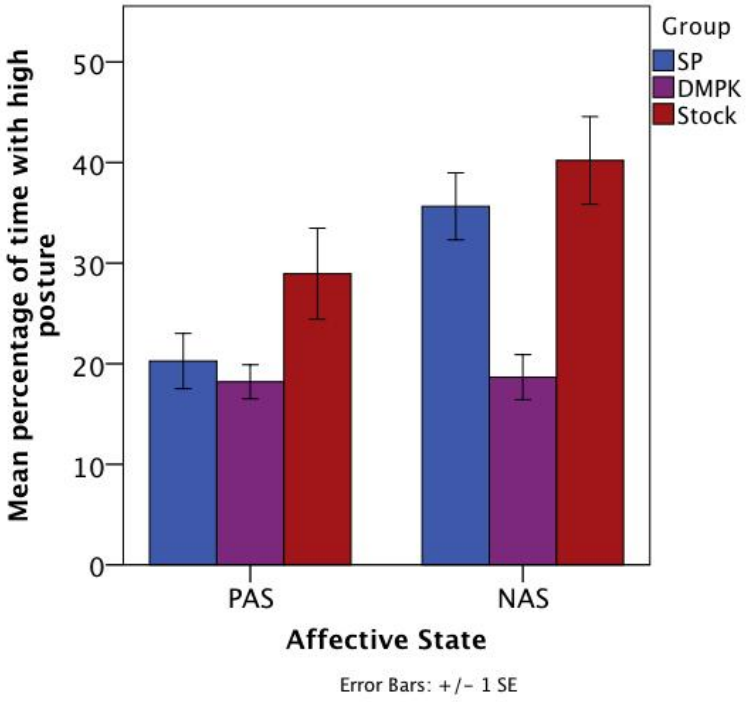


FIGURE 5.12: Mean percentage of time spent with high posture by Affective State and Group

NAS dogs spent more time with high posture than did PAS dogs. This is congruent with the finding that NAS dogs also spent more time standing alert as high posture indicates vigilance or orientation to a stimulus. The marginally non-significant interaction between group and affective state for high posture is the result of DMPK dogs showing similar levels of high posture regardless of affective state, while NAS SP and Stock dogs spend more time with high posture than do PAS dogs. This suggests that for SP and Stock dogs, but not DMPK dogs, high posture may indicates a NAS. This discrepancy may also be influenced by DMPK dogs spending less time over all with high posture than the other groups.

5.3.4.1 Number of behaviours displayed

The number of behaviours displayed was compared between affective states. NAS dogs displayed both more NWIs ($t(19)=2.762$, $p=.012$) and more total behavioural states ($t(19)=3.003$, $p=.007$). This confirms what was suggested by the comparison of total behaviours between-groups, although there is a difference in the exhibition of NWIs here. These results suggest that it is the total number rather than the type of behaviour which associates with affective state. Although this assumption was not tested, it seems likely that dogs which are engaged in more restful activities (which most PWIs are) tend not to rapidly switch behaviours and spend more time engaged in one behaviour, as suggested by Boissy et al. (2007).

5.3.5 The effects of sex

Male and female dogs differ in several aspects of development and sociality (Spotte, 2012), with males being slower to mature and females being more social and able to live closely with conspecifics. The comparison of male and female dogs within this sample will determine if there is a between-sex difference in welfare apparent, and if the behavioural indicators of welfare can be applied to both or only one sex.

When analysing the effects of sex, the SP group was excluded as the dogs were all male and exhibited behavioural differences from other groups. All male and all female DMPK and Stock dogs were compared for between-sex differences. DMPK and Stock groups were also compared to determine if the factors differing between groups (e.g. age, history of regulated procedures) had different effects on the two sexes.

As before, only behaviours occurring at more than 5% of time for at least 50% of dogs were included. This resulted in a slight change from the behaviours shown in Table 5.4, with 23 behaviours now included in analysis as shown in Table 5.7. Several

TABLE 5.7: Behaviours selected for analysis, showing means and standard deviations

Behaviour type	Behaviour	Mean % of time	SD
Location	Front of pen	49.41	20.21
	Back of pen	12.04	16.92
	Barrier	7.21	8.20
PWIs	Standing against walls	11.16	11.11
	Amicable dog	6.73	7.43
	Interact with environment	7.14	7.75
	Rest head down	12.05	17.19
	Rest head up	18.99	11.17
	Calm locomotion	5.15	6.57
NWIs	Sitting alert	21.13	12.79
	Standing alert	14.16	10.14
Postures	Tail wag (high)	19.74	15.21
	High posture	23.63	16.38
	Neutral posture	38.46	19.72
	Half-low posture	9.18	9.85
	Low posture	14.48	17.61
Others	Vocalising	6.68	10.94
	T-dog position	7.05	21.51
Events	Oral behaviours	1.88	5.84
	Paw lifts	1.61	3.86
	Circle	1.55	6.78
	All events	7.86	16,76

behaviours included here did not meet the criteria for inclusion previously due to the low occurrence in SP dogs, especially behavioural events.

These behaviours were analysed for between-sex differences, and where behaviours were normally distributed, for interactions between sex and group, or sex and affective state. The effect of sex and the interactions with group and affective state are shown in Table 5.8.

5.3.5.1 Positive and negative welfare indicators

There were three differences in behaviour between all males (n=7) and all females (n=7). Female dogs spent more time resting head up (a PWI) while male dogs spent more time standing alert (a NWI). Initially it may appear that male dogs have more negative welfare than female dogs, although it is not possible to draw firm conclusions based on only two behavioural indicators of welfare. Female dogs spent more time high tail wagging than did male dogs, however as tail wagging is a social behaviour it is unsurprising that more social female dogs spend more time tail wagging. From these data it is unclear if welfare differs by sex.

TABLE 5.8: Results of ANOVAs showing significant between-Sex differences in home pen behaviour

Effect or interaction	Behaviour	F	df	p
Sex	Rest head down	5.213	1, 63	.026
	Stand alert	5.035	1, 63	.028
	Tail wag high	5.600	1, 63	.021
	High posture	5.457	1, 63	.023
Group x Sex	Interact with environment	17.679	1, 63	<.001
	Stand alert	28.685	1, 63	<.001
	Tail wag high	4.343	1, 63	.041
	High posture	14.299	1, 63	<.001
	Neutral posture	10.327	1, 63	.002
	all events	10.217	1, 63	.002
AS x Sex	Sit alert	10.327	1, 63	.001
	Neutral posture	9.563	1, 63	.003

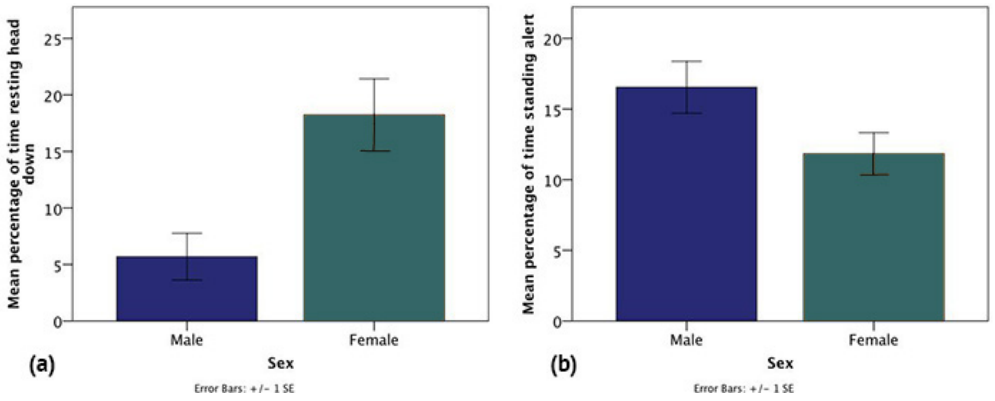


FIGURE 5.13: Mean percentage of time spent (a) resting head up and (b) standing alert in the home pen by Sex

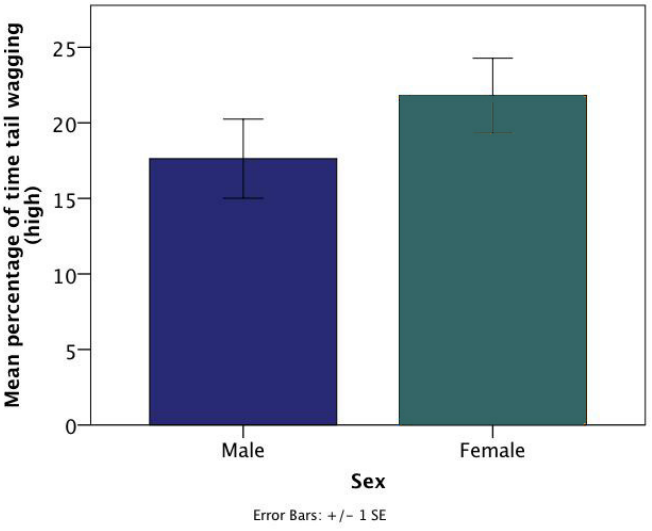


FIGURE 5.14: Mean percentage of time spent high tail wagging by Sex

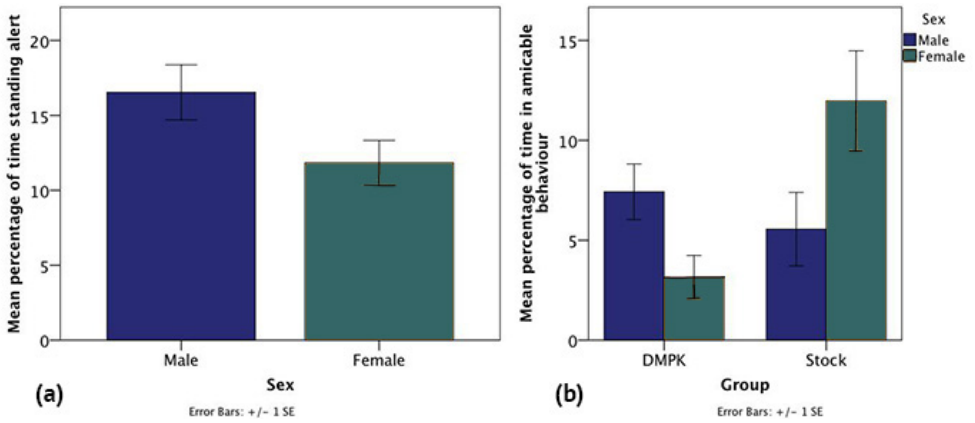


FIGURE 5.15: Mean percentage of time spent with (a) interacting with environment and (b) in amicable behaviour in the home pen by Sex and Group

5.3.6 The interactions between Group and Sex

There were more between-sex differences when DMPK and Stock groups were compared, suggesting that behaviour and perhaps welfare change for male and female dogs with age, or with changes in housing and licensed procedures.

5.3.6.1 Positive welfare indicators

For both PWIs, the opposite trend is seen between DMPK and Stock groups. Male DMPK and female Stock dogs show significantly greater levels of both interacting with the environment and amicable dog-directed behaviour. Amicable dog interactions might be expected to be higher in the more social female dogs, so that they are lower in DMPK females dogs suggests that their welfare may be worse than that of younger Stock dogs with no histories of regulated procedures.

5.3.6.2 Negative welfare indicators

Stock male and female dogs showed no difference in time spent standing against walls ($t(28)=1.433$, $p=.163$), however female DMPK dogs spent marginally more time standing against walls than did males ($t(39)=2.037$, $p=.048$). The difference between male and female DMPK dogs for sitting alert was marginally non-significant ($p=.050$), while male Stock dogs spent significantly more time sitting alert than female Stock dogs ($t(12)=2.434$, $p=.022$). The trend in NWIs is the opposite of that seen with PWIs, with male DMPK and female Stock dogs showing greater levels of PWIs, and female DMPK and male Stock dogs showing greater levels of NWIs. This further

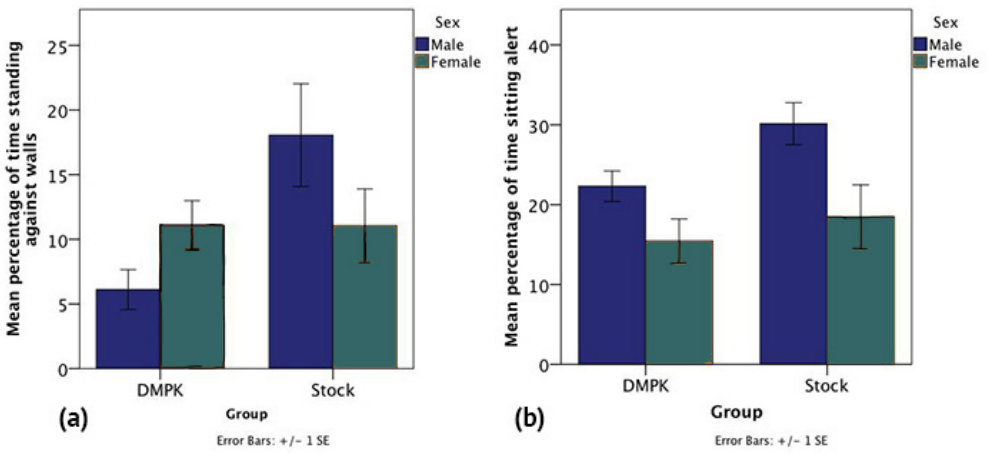


FIGURE 5.16: Mean percentage of time spent with (a) standing against walls and (b) sitting alert by Sex and by Group

supports the suggestion of welfare decreasing with increasing time in the unit for female dogs.

5.3.6.3 Posture

Differences in posture are seen between the sexes in DMPK and Stock groups. Male stock dogs spent more time with high posture than any other dogs, while female stock dogs spent more time with neutral posture than other dogs. A difference in tail wagging is seen between groups, with female DMPK and male Stock dogs spending more time tail wagging, although the difference between male and female Stock dogs was not significant. The difference in tail wagging is interesting and perhaps unexpected as female DMPK and male Stock dogs appear to be exhibiting a consistently more negative welfare state than male DMPK and female Stock dogs. However, tail wagging is not a clear indicator of welfare and so its role here is unclear.

5.3.6.4 Behavioural events

An opposite trend in the rate of behavioural events was seen between DMPK and Stock groups, with female DMPK and male Stock dogs exhibiting more behavioural events than male DMPK and female Stock dogs. All behavioural events were considered NWIs and potentially stereotypic behaviours, which supports evidence of welfare decreasing in female dogs.

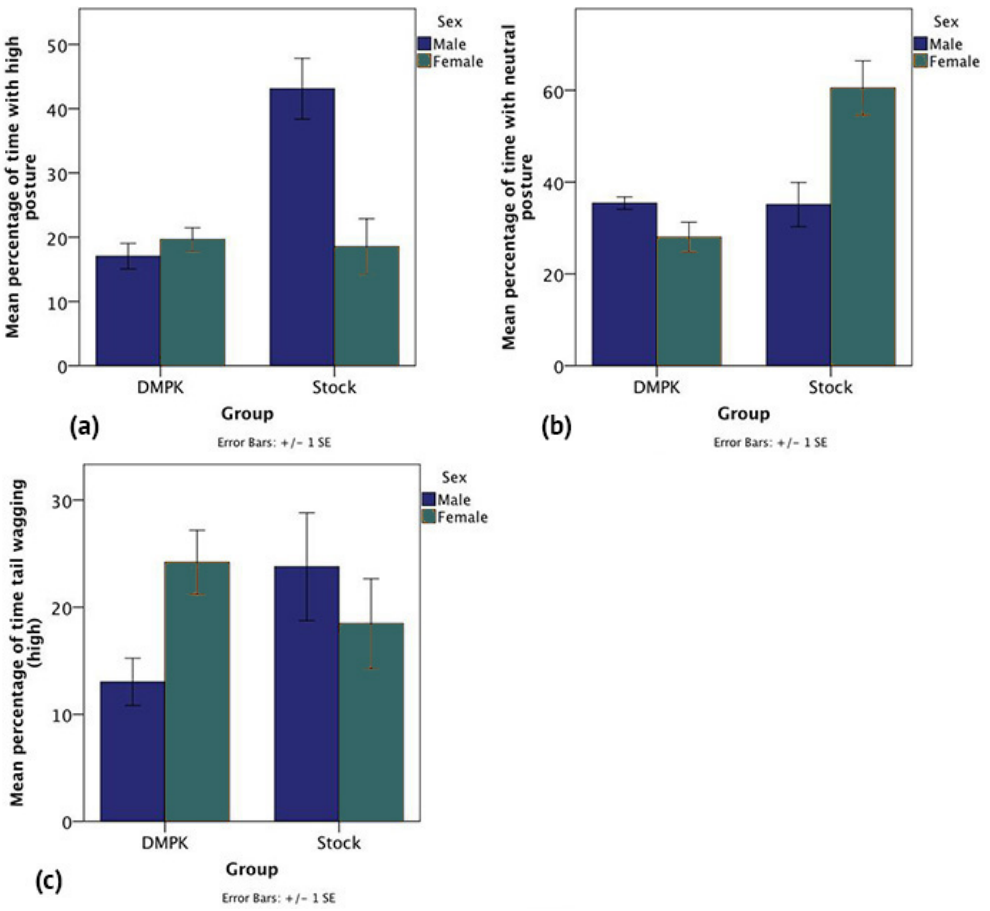


FIGURE 5.17: Mean percentage of time spent with (a) high posture, (b) neutral posture and (c) high tail wagging by Sex and Group in the home pen

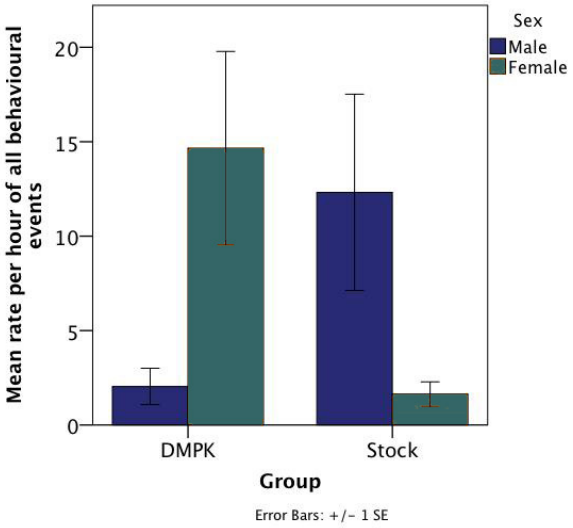


FIGURE 5.18: Mean rate per hour of behavioural events by Sex and Group in the home pen

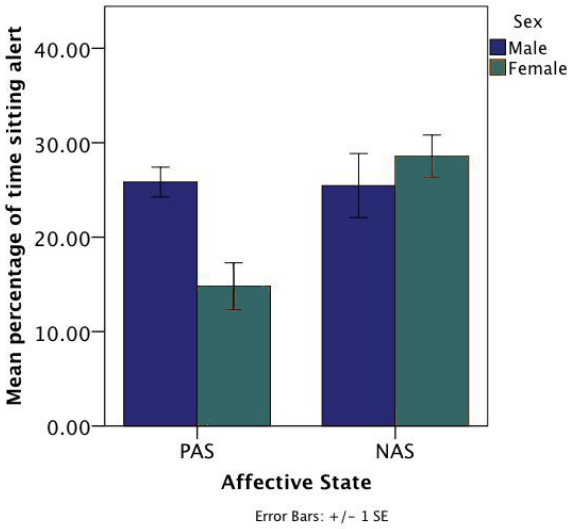


FIGURE 5.19: Mean percentage of time spent sitting alert by Sex and Affective State in the home pen

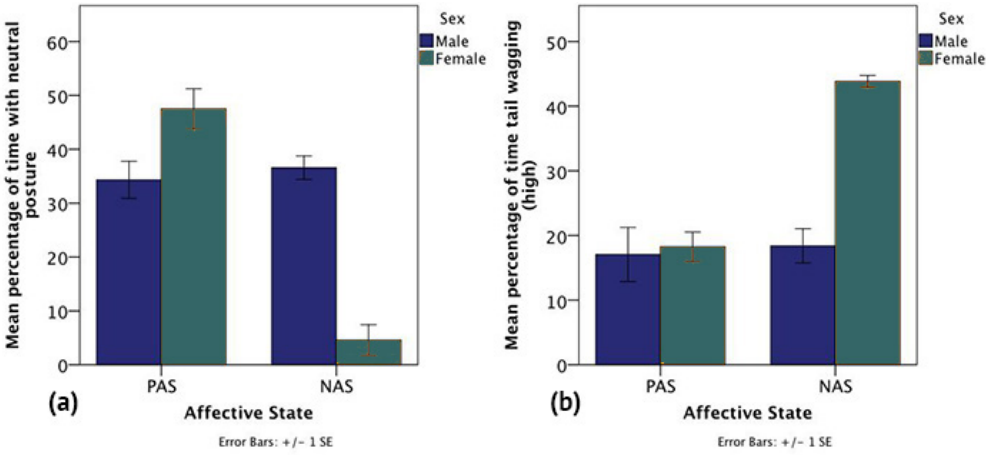


FIGURE 5.20: Mean percentage of time with (a) neutral posture and (b) high tail wagging by Sex and Group in the home pen

5.3.7 The interaction between Affective State and Sex

There were also behavioural differences between male and female dogs with differing affective states. Female PAS dogs spent less time sitting alert than other dogs, suggesting that for female dogs sitting alert may be a more sensitive indicator of affective state.

Similarly, time with neutral posture is significantly higher for PAS than NAS female dogs, while there is no significantly difference between male dogs' affective states. This again suggests that for female dogs, neutral posture may be more strongly indicative of affective state.

5.3.7.1 Number of behaviours displayed

The number of behaviours displayed was compared between sexes. Female dogs displayed more total behaviours ($t(12)=2.278$, $p=.042$) than male dogs. Since the number of total behaviours is higher in SP dogs and in NAS dogs, this suggests that female dogs have more negatively-valenced welfare than male dogs.

5.4 Results of cardiovascular analysis (SP only)

Four cardiovascular parameters were available for analysis: systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR) and corrected QT interval (QTc). Within- and between- dog stability are of particular importance in safety assessment testing (Chapter 2), variation for each of these parameters was analysed. As the PAS and NAS dogs showed significant between-groups differences for several behavioural states, parameters were also analysed to determine if this between-groups difference was also present in cardiovascular parameters. It was not possible to test for effects of sex as all SP dogs were male.

5.4.1 Within-dog variation

For all dogs (n=8), there was no significant effect of observation session (10) on any of the four parameters ($.238<p<.962$). There was also no significant interaction between observation session and affective state for any parameter ($.787<p<.996$). Within-dog variation is therefore considered to be low at baseline in the home pen, regardless of affective state.

5.4.2 Between-dog variation

QTc was calculated for five dogs as correction factors were unavailable for the others. Two of the dogs (Bob and Nibbler) had QTc values which were both significantly different from the others' and considerably outwith the normal range and so had to be discounted. The values recorded were incompatible with normal heart function and therefore most likely due to analysis errors by the Notocord software. This resulted QTc being available for only three dogs and it was no longer possible to make meaningful between-dog comparisons; QTc was therefore discounted from further analyses.

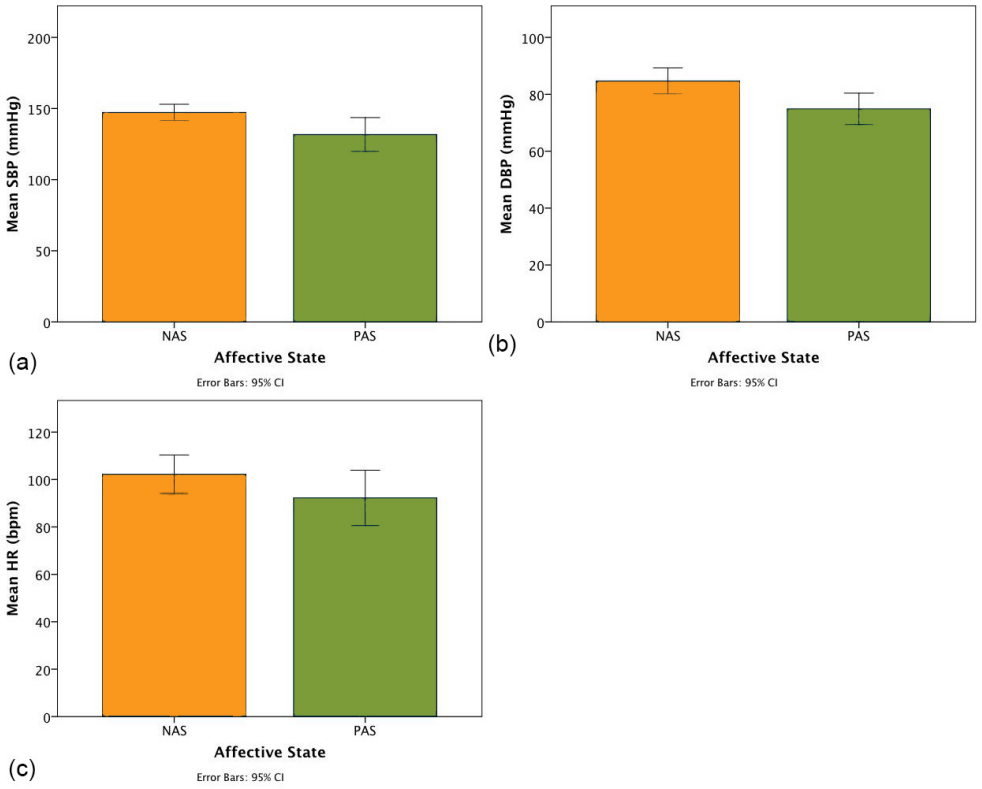


FIGURE 5.21: Mean (a) SBP, (b) DBP and (c) HR by Affective State at baseline

In contrast, for two of the three parameters, there was significant between-dog variation (Table 5.9). Heart rate showed a non-significant trend towards between-dog variation. In the absence of within-dog variation, this is not concerning as within-dog comparisons are commonly used, unless the difference is due to systematic variation or a confound which is otherwise unaccounted for.

TABLE 5.9: Results of ANOVAs to determine between-dog variation in four CV parameters

Parameter	**F(7, 64)**	**p**
SBP	14.90	<.001
DBP	2.42	.030
HR	1.93	.081 (NS)

5.4.3 The effects of Affective State

Several important behavioural welfare indicators varied by affective state, so it was important to determine if this difference extended to cardiovascular parameters. To this end, the blood pressures and heart rates of PAS and NAS dogs were compared.

Table 5.10 shows the results of the independent-samples t-tests performed for the three cardiovascular parameters.

As can be seen in Figure 5.21 (a, b), the significant difference between PAS and NAS dogs for SBP and DBP is a result of higher values for the NAS group. Heart rate (c) appears to be showing a similar trend, however within-dog variation was greater for heart rate than for blood pressure.

TABLE 5.10: Results of independent-samples t-tests to determine effects of AS on three CV parameters at baseline

Parameter	**t(3)**	**p**
SBP	6.68	.014
DBP	7.93	.006
HR	2.15	.147

As the magnitude of the effect of welfare is the parameter of interest, 95% confidence intervals, mean and Cohen's d effect size is presented in Table 5.11.

Note that for SBP and DBP, the 95% confidence intervals overlap by only <2mmHg and <0.2mmHg respectively, suggesting that the PAS and NAS groups are distinct. The distinction is less clear for HR, with PAS dogs having an upper limit which overlaps the lower limit of the NAS group.

Using the formula

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s.d.} \tag{5.1}$$

the following effect sizes were calculated: SBP: 0.60; DBP: 0.66; HR: 0.36, which are considered medium, medium and small effect sizes respectively (Cohen, 1992).

TABLE 5.11: Confidence intervals and means for CV parameters by Affective State

	PAS		NAS	
	Mean	95% CI	Mean	95% CI
SBP	131.72 mmHg	119.81- 143.63 mmHg	147.25 mmHg	141.51-153.00 mmHg
DBP	74.86 mmHg	69.32-80.40 mmHg	84.70 mmHg	80.12-89.27 mmHg
HR	92.24 bpm	80.56-103.93 bpm	102.23 bpm	94.11-110.36

5.5 General Discussion

The evaluation of baseline behaviour and cardiovascular function in the home pen has identified a number of behaviours that can be easily identified and observed, as well as identifying key differences between groups, between affective states and between sexes. These differences elucidate the relationship of these behaviours to the welfare state of the dog, and allow inferences to be made about the factors which may influence welfare.

5.5.1 Behaviours selected for inclusion

Of the behaviours included in the scoring scheme for baseline observations (see Appendix B), 18 met the criteria for inclusion in analyses of all dogs. This expanded to 23 behaviours when analysing only the behaviour of the DMPK and Stock groups, most likely due to some behaviours in the SP group being under-represented, caused by lower-resolution video footage and the lack of any activity in the housing zone during observations.

Many of the behaviours measured here at baseline in the home pen are 'static' measures, i.e. they do not reflect a dog's response to its environment, and it is not until 'dynamic' observation of behaviours takes place in response to behavioural challenges (Chapter 6) that their indication of positive or negative welfare will become apparent. For example, location at the front or the back of the pen does not appear to indicate a positive or negative change in welfare and baseline measurement of location has not provided information about welfare. However, a change in a dog's location in response to a behavioural challenge may well provide information about its reaction.

Similarly, few behavioural events occurred at a sufficient rate to meet the criteria for inclusion in analysis. Many of these events are suggested to be a response to a stressor in the environment (see Appendix B), so it is not surprising that few were seen at baseline; their appearance following a behavioural challenge would provide information about a dog's response to the challenge (see Chapter 6). There were however several key differences in behaviour at baseline which do indicate differing welfare states.

5.5.2 Between-groups differences in behaviour

Many of the between-groups differences in behaviour provide information about the environment in which the dogs are housed, rather than specifically about welfare. For example , SP dogs spend more time at the back of the pen, and Stock dogs spend more time at the front of the pen. There is however no evidence that SP dogs are more

restful or that Stock dogs are more alert or vigilant than other groups. The differences are likely to reflect the environment and the age of the dogs; SP dogs were housed in an area which was restricted during recording, while Stock dogs were housed in an area with many other dogs and frequent staff activity (DMPK dogs experiencing an intermediate level of activity) and so their location in a part of the pen that provided or prevented visual contact with the rest of the housing zone is likely to be reflective of this.

Similarly, standing against walls was highest in Stock dogs and lowest in SP dogs; a behaviour which reflects excitement or agitation but despite being most frequently being exhibited by Stock dogs (a predominantly PAS group), this behaviour was more frequently seen in NAS than PAS dogs. This suggests that again the difference in behaviour is explained by the considerable difference in activity levels in the housing zone; the age of the dogs is also likely to be a factor as young dogs were more frequently observed to be excitable in response to the presence of staff without other signs of a negative affective state or agitation.

The differences between DMPK dogs and the other groups perhaps provide more information about their welfare. DMPK dogs spent less time resting head down (a PWI) but this does not seem to be reflected in an increase in alert or agitated behaviour. This is seen also in DMPK dogs spending less time with high posture than other groups. DMPK dogs also spend less time with neutral posture than other groups. These taken together suggest that DMPK dogs were neither sleeping (resting head down) or alert (or with high posture) but engaged in a variety of intermediate activities and behaviours.

5.5.3 The influence of Affective State

The differences in behaviours seen between PAS and NAS dogs are more consistent and distinct than the differences between groups and provides clear indications of differences in welfare; this was expected as the literature suggests that a negative affective state and good welfare are incompatible. PAS dogs spend less time at the barrier, a measure of agitation as it reflects movement between pens; less time alert, a measure of vigilance that in high levels reflects an inability to relax in the environment and less time with high posture, also a measure of vigilance. As may be expected from this pattern of behaviour, PAS dogs also spend more time resting head up, a behaviour which indicates that dogs are sitting or lying calmly but not asleep.

The predominant distinction between PAS and NAS dogs appears to be best described as a degree of agitation: NAS dogs spend more time moving between pens, alert and

with high posture (vigilant) and less time resting calmly (head up). This may also explain why NAS dogs have higher blood pressure in the home pen than PAS dogs.

One incongruent finding is that the relationship between AS and high posture is reversed for DMPK dogs, with PAS dogs exhibiting more high posture. DMPK dogs exhibit less high posture as a group than other groups; as previously discussed they also exhibited less neutral posture, with the likely cause being a greater variety of postures exhibited. Since there are no other instances of a NWI being associated with a PAS in DMPK, it may be that high posture was associated with a greater variety of behaviours resulting in this discrepancy.

Although affective states are less easy to identify than groups such as SP, DMPK and Stock, differences between affective state provide more information about welfare and better identify individuals with negative welfare. A clear pattern of behaviours as seen here provides an easy-to-use tool to identify those individuals and allows identification of affective states without the need for measurement by cognitive bias testing or mechanical pressure threshold testing.

5.5.4 Between-sex differences

The effects of sex on welfare are less clear than other factors and as such it is not possible to draw strong conclusions about differences in welfare between the sexes. Female dogs exhibited more signs of relaxed behaviour (e.g. resting head up) while male dogs were more vigilant (standing alert). One interesting finding was that the opposite patterns were seen in DMPK and Stock dogs, suggesting that welfare may change over time differently for male and female dogs. Male DMPK and female Stock dogs showed better welfare than female DMPK and male Stock dogs. As there were only seven dogs of each sex it is difficult to draw conclusions about the exact nature of any welfare differences, but given the differences in sociality in male and female dogs, it is possible that this is a factor. The structure of canine societies means that young male dogs in large groups may encounter more conflict than older dogs in smaller, more stable groups, while young female dogs living in large groups are more likely to co-exist peacefully (Spotte, 2012).

5.5.5 Number of behaviours displayed as a welfare indicator

The number of behaviours displayed (PWIs, NWIs and total behaviours) was calculated. A greater number of PWIs and lower number of NWIs should be an indication of positive welfare. SP dogs and NAS dogs showed a greater number of both

PWIs and total behaviours, in contrast to what was expected as a greater variety of PWIs should logically indicate better welfare. However, this is consistent with Rutherford's use of fractal analysis in broiler chicken behaviour, in which an increasing number of transitions between behaviour is associated with decreasing welfare.

Given the patterns of behaviours seen in PAS such as resting head up or head down or interacting with the environment, it is possible that PAS dogs remain engaged in one calm behaviour for a period of time while NAS dog engage in a rapid succession of behaviours including both PWIs and NWIs. This is reflected in the behaviours exhibited by NAS dogs such as rapid locomotion, standing alert and moving between pens. These results suggest that the type of behaviours exhibited may be more important than simply the number.

5.5.6 The relationship between behaviour and cardiovascular function

Much like behaviour in the home pen at 'baseline', cardiovascular function at baseline cannot provide information about responses to positive or negative events. The results of this study have however shown a difference in blood pressure in home pen in the differing welfare types. The finding that NAS dogs have higher blood pressure (both SBP and DBP) is of particular concern to quality of data output. Disregarding the direct measure of blood pressure in safety assessment, even a small increase in blood pressure is known to have health consequences (Lewington et al., 2002), so dogs exhibiting a NAS may respond differently to health challenges or toxicity testing. The analysis of 95% CIs for blood pressure showed that the two affective states are almost distinct, meaning that they should not be considered one population for the purpose of safety assessment, although the use of factorial designs could incorporate this. A medium effect size was detected using Cohen's d, which at baseline is greater than would be expected for a population of dogs with near-identical life experiences.

While SP dogs act as their own controls within studies, it is still of concern to have one sample of dogs with a higher blood pressure value at baseline as any increase also decreases the sensitivity of measurement, meaning that subtle changes in cardiac function may be more difficult to detect and that larger samples sizes are needed to detect such effects. This same problem may lead to a Type II error, the lack of detection of an effect where one is present. Unwanted or unattributed variation leading to increased sample sizes is contrary to the principles of good scientific practice or humane experimental procedure.

Given the consistent pattern of behaviour associated with affective states across SP and other dogs (high posture, vigilance, stereotypic behaviours), it is highly likely that

this difference in blood pressures is also present in dogs which do not have implanted telemetry. At the very least it seems likely that dogs subject to long-term use such as SP dogs would show the same patterns which would remain undetected. Unlike blood pressure, heart rate showed no significant differences relating to welfare, although there was a trend towards heart rate being higher in NAS dogs. The 95% CIs showed that while PAS dogs had a much lower limit of the CI, the upper limit overlapped with NAS dogs. This suggests that the mechanism which causes the increase in blood pressure seen in NAS dogs does not also act on heart rate at baseline, and as heart rate is influenced by many factors such as anticipation and activity, baseline heart rate is not a clear indicator of heart rate.

Unfortunately, insufficient data were available to examine the effects of welfare on QT interval. As an important parameter in the safety testing of compounds destined for human use, any unwanted variation would be of great concern. Given the variation in blood pressure identified here, it is important to discover if there is an effect of welfare on cardiac parameters and this should be considered in future data collection.

5.5.7 Final conclusions

The investigation of behaviour and cardiovascular function in the home pen was conducted following an assessment of affective state in which two clear patterns of affective state were found: PAS and NAS. One of the most important findings of this study has to be that even in the home pen with no stimulus, NAS dogs have higher blood pressure than PAS dogs. The investigation of cardiovascular function in the following chapter will be crucial to evaluating the effects of this on quality of data output as any difference in baseline function may well influence response to challenges. Differing responses to challenges would be of concern as it would suggest that different data output may be obtained in response to test compounds.

Comparisons of the three Groups and between-sex suggests that welfare is greatest in naive Stock dogs and most negatively-valenced in the most intensively-used SP dogs. DMPK dogs, which are long-term but less intensively used have intermediate welfare. Welfare may also change over time by sex, with male dogs coping better with long-term use than female dogs however further investigation with more varied samples are needed to confirm this.

As with cardiovascular data, the most striking comparisons were between Affective States. PAS dogs show contrasting patterns of welfare to NAS. The identification of positive and negative affect in Chapter 4 suggested that these groups of dogs would exhibit different patterns of behaviour and this study has confirmed this. Further, the

behaviours associated with PAS and NAS are those which were *a priori* expected to be associated with positive and negative welfare respectively. For example vigilant behaviours such as sitting or standing alert, having high posture or moving frequently between pens were unlikely to be reflective of positive welfare.

This study suggests that the behaviours most strongly associated with negative welfare are behaviours relating to an increased vigilance or activity in the home pen in the absence of stimulation, while restful or exploratory behaviours are associated with positive welfare. The next study will evaluate changes in behaviour in response to behavioural challenges to determine which are most sensitive to changing welfare and therefore most suitable for inclusion in the Welfare Assessment Framework.

CHAPTER 6

Assessing welfare and data quality in response to behavioural challenges

Developing the Welfare Assessment Framework III

In Chapter 5, three groups of dogs (SP, DMPK and Stock) were observed in the home pen. Data on affective state and mechanical pressure threshold (MPT) sensitivity were collected (Chapter 4). A consistent pattern of behaviour, cardiovascular parameters, affective state and mechanical pressure threshold was found, distinguishing between those with positive welfare and those with negative welfare. One of the overarching aims of this project was to develop a Framework in the form of an easy-to-use tool for monitoring welfare, which means that the parameters chosen much be sensitive to changing welfare. To ensure this, the same groups of dogs are subject to behavioural challenges, both positive and negative in nature, using the same behavioural measures used in Chapter 5 to identify the most suitable behaviours for inclusion in the Framework. As in Chapter 5, cardiovascular data were also available for SP Dogs, allowing an examination of the effects of the behavioural challenges on the quality of data obtained from the dogs. The result of integrating data from Chapters 4-6, the Welfare Assessment Framework, is also presented.

6.1 Background

From what is known about dogs (Chapter 1) and the laboratory environment (Chapters 1 and 2), there are several very common events in the life cycle of a dog

which have the potential to positively- or negatively-influence welfare. The paucity of data relating to housing and husbandry practices means we have little information on which to base best practice. Human interaction (Section 1.2.1.) appears to be very important to dogs, due to the selective breeding of traits which predispose them to cooperate with us and seek contact with us. Human contact is also an unavoidable aspect of the laboratory environment, with husbandry and regulated procedures carried out by one or more members of staff. This means that there are many points in the life cycle of the dog during which human contact can have a positive or negative impact. Brief periods of human interaction (unstructured, positive interactions) have been shown to have considerable positive impacts on behaviour and physiology in the dog (see Table 1.7 in Section 1.2.1.), from heart rate, blood pressure, cortisol, immune factors and oxytocin. For these reasons, a period of human interaction (HI) was chosen as a positive (social) challenge.

Conversely, social isolation has been shown to have a considerable negative impact on dog welfare and physiology, depending on the severity and duration of the isolation. Single-housing from pen mates (although not true isolation) is a regular occurrence in the life cycle of the dog, whether for veterinary reasons, measurement of food consumption, or to prevent cross-contamination of test compounds, or corprophagy. While dogs may maintain visual, auditory and/or olfactory contact during single-housing, there is no physical contact between dogs. Although it is recommended that dogs separated from pen mates are given regular human contact, this is often not practised due to time constraints in the laboratory environment. Although not in force at the time of this study, European Directive 63/2010/EU has since come into force, and under this legislation, single-housing becomes a regulated procedure at greater than four hours, while "prolonged" isolation is considered a substantial severity procedure in the dog. A one-hour single-housing period was a common occurrence for dogs in this facility for a variety of reasons and so this was chosen as a negative (social) challenge.

Feeding in the laboratory is usually a standardised protocol, with the same weight of diet presented every day. Often, enrichment using food is overlooked or discounted because of a perceived cost to data quality. However, one of the most commonly-cited enrichments for laboratory-housed (e.g. Schipper et al., 2008) or shelter-housed (e.g. Wells, 2004a) dogs is the provision of toys, in particular food toys (see Chapter 2 for review). These toys provide an interactive object which dispenses a pleasant treat, increasing calm activity and decreasing indicators of negative welfare in the home pen. These food toys also have the added benefit of requiring little input from staff, making them practical in the laboratory environment. A food toy (FT) placed in the home pen was chosen as a positive (physical) challenge.

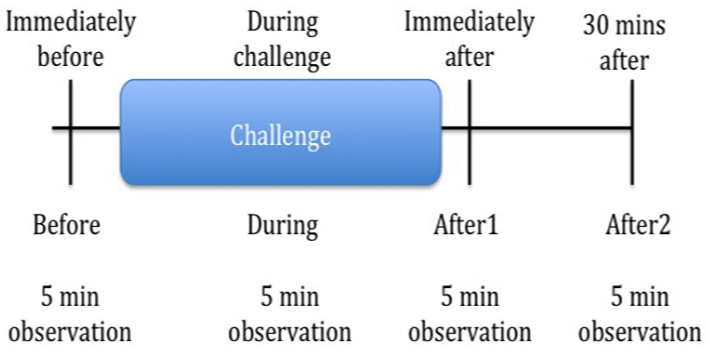


FIGURE 6.1: Time points selected for analysis for all Challenges

One of the most common negative experiences for a laboratory-housed dog is restraint before a procedure. While restraint is also performed for other reasons (health checks, transport), it is also an (unreliable) signal that a regulated procedure is about to take place. Restraint also removes the element of control that the dog has over its environment and experiences and so increases the aversiveness of a procedure. Restraint is not an intrinsically aversive experience for a dog, but these factors often mean it is. Because of the necessity of close contact between dog and handler during restraint, there is a social aspect to it, however in this instance it was chosen as a predominately negative (physical) challenge.

6.2 Methodology

6.2.1 Dogs

The same population of dogs were used for this study as for the analysis of baseline data in Chapter 5 (see Chapter 3 for demographic details). In all cases, data collection for this study immediately followed data collection for the previous study.

6.2.2 Methods

Four behavioural challenges were therefore conducted: human interaction, single-housing, feeding toy and restraint. This resulted in two challenges being of a physical nature and two of a social nature (see Table 6.1). All challenges followed the same protocol for selection of time points for analysis.

Table 6.1: Nature of Challenges

Type	**Social**	**Physical**
Positive	Human Interaction	Feeding toy
Negative	Single-housing	Restraint

Table 6.2: Duration and location of the four challenges

Challenge	**Duration**	**Location**
Human Interaction	10 min	Home pen (SP)
		Play area (DMPK and Stock)
Single-housing	1 hour	Home pen
Feeding Toy	10 min	Home pen
Restraint	5 min	Procedure pod

The four challenges were different in nature, Table 6.2 shows the duration and location of each of the challenges. The human interaction and feeding toy challenges were each of 10 minute duration, as 10 minutes interaction per pen was determined to be practical from a staff perspective, and the feeding toy would last no longer than 10 minutes. It was necessary to conduct the human interaction challenge in the home pen for SP dogs in order to capture cardiovascular data, however this was changed to the indoor play areas for DMPK and Stock dogs because of the greater space and presence of play equipment. The restraint challenge was conducted in the procedure pod closest to each dog's home pen, with the five minute duration corresponding to the length of time for which a dog would be on the table during examinations or procedures. The length of the single-housing challenge was chosen to be of sufficient length to mimic common husbandry procedures, without causing distress to the dogs; dogs were commonly isolated for feeding, before or after dosing or during other husbandry activities for durations of more than one hour.

Full details of data analysis can be found in Chapter 3. The same methodology was adopted for recording behaviour and cardiovascular data as in Chapter 5, with each sample being of five minutes duration. For the purposes of analysis, behavioural states occurring for less than 5% of time for 50% of dogs and events occurring at less than two per hour for 50% of dogs were discounted as these would not be behaviours which could be easily observed and would not be suitable for use in a Welfare Assessment Framework.

Data were tested for normality and normally-distributed data were analysed using a factorial, repeated measures ANOVA, with a within-subjects factor of time and two between-subjects factors of Group (SP, DMPK and Stock) and Affective State (PAS or NAS). Where significant interactions were found, the appropriate post-hoc test was

used to determine the direction of the interaction. Non-normally distributed data were analysed using the appropriate non-parametric test (related-samples Friedman's two-way analysis of variance by ranks, independent samples Mann-Whitney U test or independent samples Kruskal-Wallis test).

In Chapter 5, medium effect sizes were detected in CV data using Cohen's d when investigating the effect of affective state at baseline in the home pen. In this Chapter, the time points of interest are during and immediately after the challenges, as it is already apparent that there are differences present at baseline. The magnitude of the effect of affective state on CV responses to challenges will elucidate whether welfare has interacted with stimuli which are common in the environment to increase or decrease variance at the group or individual level.

6.3 Results

For the four challenges, there were a number of effects of time, of Group and of Affective State on behaviour and cardiovascular parameters. This section summarises the key findings and a full breakdown of analysis can be found in Appendix D. Tables 6.3 -6.5 summarise the changes in behaviour seen for all dogs, by Group and by Affective State; changes are coded as positive, negative or **context-dependent** increases (↑) or decreases (↓).

6.3.1 Summary of responses to Challenges

6.3.1.1 Response to challenges for all dogs (n=22)

Table 6.3 shows the effects of each behavioural challenge on the behaviour of *all dogs*. As might be expected, the two positive challenges resulted in positive changes and the negative challenges resulted in more negative changes; supporting the classification of these events.

Human interaction had an effect on the greatest number of behaviours, most of which were positive, however significant increases in heart rate and diastolic blood pressure occurred during the interaction period. Conversely, single-housing had few effects on behaviour, although these were predominantly negative; and no effects were observed on cardiovascular parameters.

Similarly, the feeding toy had positive effects on behaviour and no behavioural indicators of welfare increased in duration as a result of the feeding toy challenge.

TABLE 6.3: Combined immediate and longer-lasting effects of challenges on all dogs (n=22)

	Human Interaction	Single-housing	Feeding Toy	Restraint
Behaviour				
Barrier				↑
Resting head up	↓		↓↑	↓
Resting head down	↓		↓	
Interacting with environment		↓		↑
Sitting alert	↓	↑	↓	
Standing alert	↓			
Calm locomotion		↑		
Play behaviours			↑	
High tail wagging	↑		↑	
Low tail wagging	↑			
High posture	↓		↓	
Neutral posture	↑	↓	↑	
Half-low posture	↓	↑		
Crouch/tremble				↑
Struggling				↑
Panting			↑	
Behavioural events				↑
HR (n=7)	↑			
SBP (n=7)				↑
DBP (n=7)	↑			↑

Although resting behaviours decreased, no corresponding increase in alert behaviours occurred, with dogs spending more time engaged in interactive of playful behaviours. The most striking pattern of negative welfare indicators can be seen in response to the restraint challenge; positive welfare indicators decreased and negative welfare indicators increased. Interaction with the environment increased following restraint, but this is likely to be a response to removal from the home pen rather than an indication of positive welfare. Increases in blood pressure also occurred during restraint which returned to baseline 30 minutes following return to the home pen; the increase in heart rate was marginally non-significant.

Restraint caused increases in NWIs, SBP and DBP. The immediate behavioural changes seen during or after the challenges were rarely different from the behavioural changes seen 30 minutes after the challenge, with the exception of resting head up, which decreased during the Feeding Toy challenge, but had significantly increased over the percentage of time seen before the challenge.

Fewer of the behavioural changes seen in Table 6.3 are evident 30 minutes after the challenge. Both human interaction and the feeding toy resulted in several positive changes in behaviour (posture, resting behaviour) which were long-lasting. Single-housing had little long-lasting effect on behaviour, with the exception of interacting with the environment which decreased. Restraint caused dogs to be more active 30 minutes later, as seen in the decrease in resting behaviour and increased time at the barrier and interacting with environment.

6.3.1.2 Response to challenges by Group

The life histories of the dogs meant that Group was likely to influence responses to challenges. Staff contact, periods of single-housing, access to novel toys and regulated procedures all varied between the groups (Chapter 3), so the association with these stimuli was likely to be different. Table 6.4 shows the effects of each challenge on the behaviour of each group. Stock dogs showed the fewest changes in behaviour in each of the challenges, while SP and DMPK dogs showed a greater number of behavioural changes in response to the challenges.

Human interaction had a much greater effect on SP dogs than the other groups, with a number of positive and negative changes. The increases in high posture, half-low posture, alert and tail wagging behaviour suggest a higher level of excitement in SP dogs following the interaction period; in contrast, single-housing had little impact on SP or Stock dogs, while DMPK dogs showed several negative changes in behaviour. The differences in response to these challenges is a reflection of the experience of the dogs; SP dogs experienced little human contact and became very excited by staff presence, while DMPK experienced regular positive staff contact and single-housing was associated with regulated procedures.

Feeding toy and human interaction both had a singular positive change in behaviour, while restraint was the only challenge to cause a negative change (increased rate of behavioural events). SP and DMPK dogs both showed greater responses (both positive and negative) to all challenges, although SP dogs tended to show a greater change in behaviours with the exception of single-housing. SP dogs showed little response to single-housing in comparison to DMPK dogs, while DMPK dogs showed a decreased response to restraint compared to SP dogs. SP dogs showed an increase in a number of positive and negative behaviours as a result of the human interaction challenge, while both SP and DMPK showed a small number of positive and negative changes in behaviour as a result of the feeding toy challenge.

TABLE 6.4: Effects of challenges on dogs by Group

Challenge	**Human Interaction**			**Single-housing**			**Feeding Toy**			**Restraint**		
Group	SP	DMPK	Stock	SP	DMPK	Stock	SP	DMPK	Stock	SP	DMPK	Stock
Behaviour												
Location												
Front	↑	X	X	↓	X	X	↑	X	X			
Rear				↑	X	X				↑	X	X
Barrier							↑	X	X	↑	X	X
Positive welfare indicators												
Interacting with environment	↑	X	↑	↑	X	X						
Resting head down				X	↑	X						
Play	↑	X	X				↑	X	X			
Calm locomotion	X	↓	X									
Amicable behaviour										↓	X	X
Negative welfare indicators												
Standing against walls							↑	X	↑			
Standing alert	↑	X	X									
Sitting alert							X	↓	↓			
Struggling										↑	X	X
Posture												
High posture	↑	X	X				↑	X	X			
Neutral posture	↑	↑	X	X	↓	X				↓	X	X
Half-low posture	↑	X	X	X	↑	X						
Low posture										↑	↑	X
High tail wagging	↑	X	X									X
Low tail wagging										X	↑	X
Behavioural Events												
Paw lifts				X	↑	X						
Panting							X	↑	↑			
All behavioural events				X	↑	X	X	↓	X	↑	↓	↑

6.3.1.3 Response to challenges by Affective State

While all dogs, and between-groups analysis has provided some information for the Welfare Assessment Framework, it is the comparison of the two affective states which shows the clearest effect and strongest behavioural indicators of welfare. Consulting Table 6.5, it can be seen that all but one change in behaviour for NAS dogs was a negative change; this was an increase in negative welfare indicators rather than a decrease in positive welfare indicators. Of all the behaviours which were influenced by affective state, only one changed for PAS, but not NAS, dogs: resting head down following restraint. This confirms what was apparent in the literature, that it is often more easy to identify indicators of negative welfare, than of positive welfare.

Table 6.5 shows the changes in behaviour caused by each challenge by AS. It is apparent that NAS dogs are often showing changes in behaviour not seen in PAS dogs, other than in the instance of increasing time resting head down (restraint). While all dogs showed changes in behaviour as an immediate response to challenges (Table 6.3) NAS dogs often showed a greater (and more strongly negative) response which often persisted beyond the time point immediately after the challenge. This suggests lower within-subject stability for the NAS dogs, and well as greater between-subject variation when comparing PAS and NAS dogs.

Figures 6.2-6.5 present a diagrammatic depiction of the significant results of the four challenges. As before, negative and positive changes in behaviours are shown, with increases (↑) and decreases (↓) depicted with arrows. Some changes applied to PAS or NAS dogs only. Significance is shown at the $<.05$ level (*) and $<.001$ (**) level.

TABLE 6.5: Effects of challenges on dogs by Affective State

Cognitive Bias	PAS (n=14)	NAS (n=8)
Behaviours		
Human interaction		
Half-low posture		↑
Oral behaviours		↑
Paw lifts		↑
Stereotypies		↑
Single-housing		
Panting		↑
Feeding Toy		
Interacting with environment		↑
Restraint Resting head down	↑	
Standing alert		↑
Struggling		↑
Stereotypies		↑
HR		↑
SBP		↑
DBP		↑

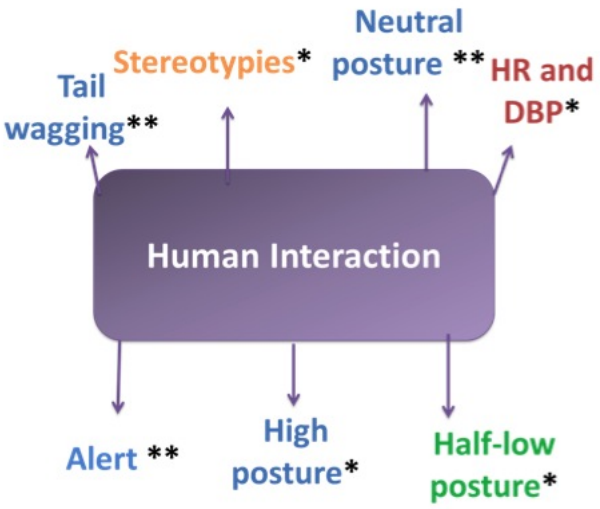


FIGURE 6.2: Diagrammatic representation of the effects of human interaction, showing the effects of Affective State

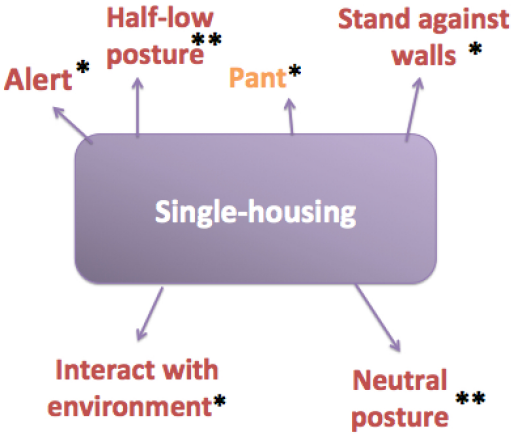


FIGURE 6.3: Diagrammatic representation of the effects of the feeding toy, showing the effects of Affective State

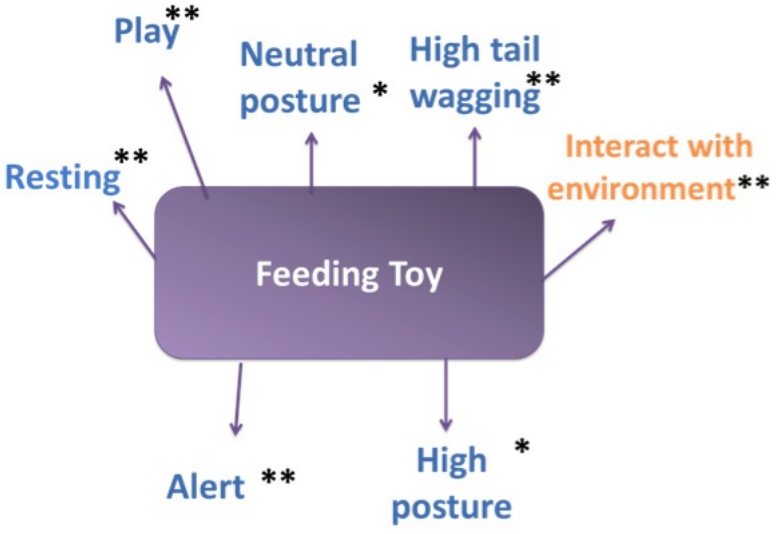


FIGURE 6.4: Diagrammatic representation of the effects of the feeding toy, showing the effects of Affective State

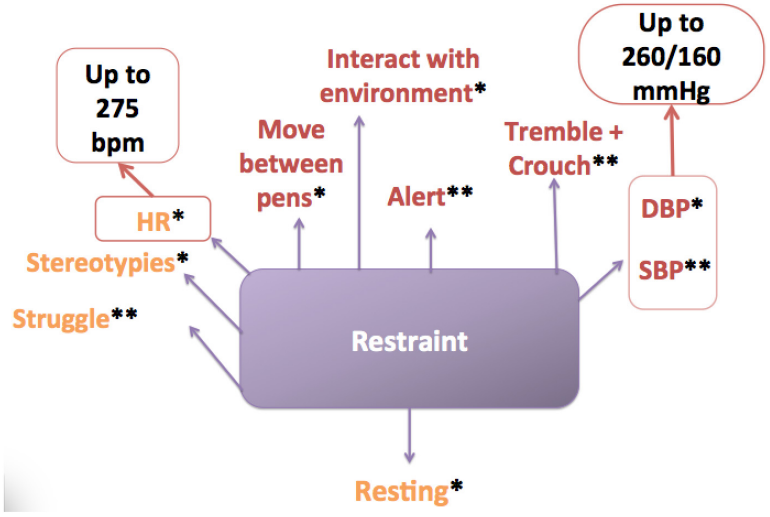


FIGURE 6.5: Diagrammatic representation of the effects of restraint, showing the effects of Affective State

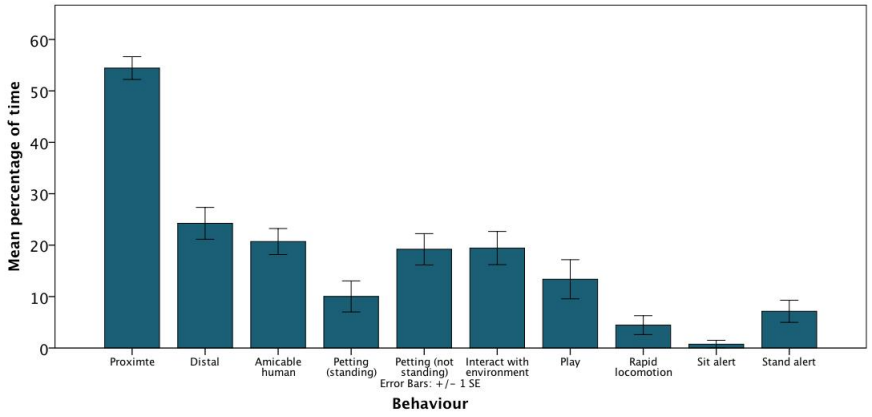


FIGURE 6.6: Behaviour during Human Interaction

6.3.2 Behaviour during Challenges

The key behaviours exhibited during each of the challenges are presented below. Many of these behaviours were not seen in the home pen during baseline data collection, highlighting the need to examine behaviours as responses to events rather than in the absence of stimuli. The telemetry implant of one SP dog (Nibbler) ceased functioning between the collection of baseline data and the beginning of this study, meaning that data were available for seven dogs only. Behavioural data continued to be collected for Nibbler.

6.3.2.1 Human Interaction

Dogs spent more than 50% of time proximate to the experimenter. Together, amicable behaviour, petting and petting while standing on the experimenter made up more than 60% of time. Petting without standing was approximately twice as common as petting with standing. Other desirable behaviours such as interacting with the environment (19.43%) and play (13.37%) were also common behaviours. Undesirable behaviours such as rapid locomotion and alert behaviours accounted for less than 13% of time.

Neutral posture was displayed for more than 50% of time, with high posture present for around 30% and half-low posture for 8%. High tail wagging was displayed for almost 80% of time, with low tail wagging present for 12%. The mean rate per hour of behavioural events was 10.

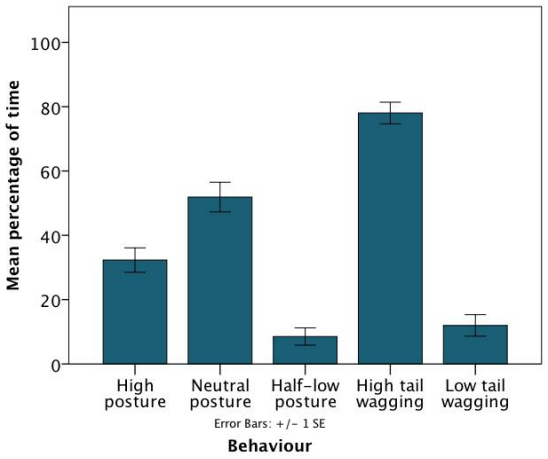


FIGURE 6.7: Posture during Human Interaction

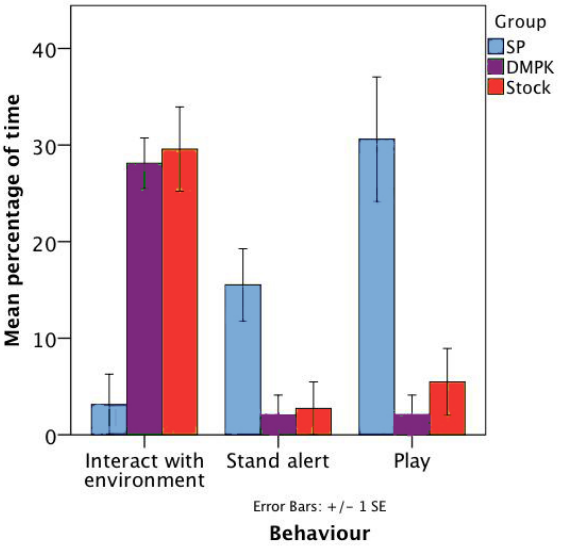


FIGURE 6.8: Mean percentage of time interacting with environment, standing alert and playing during HI by Group

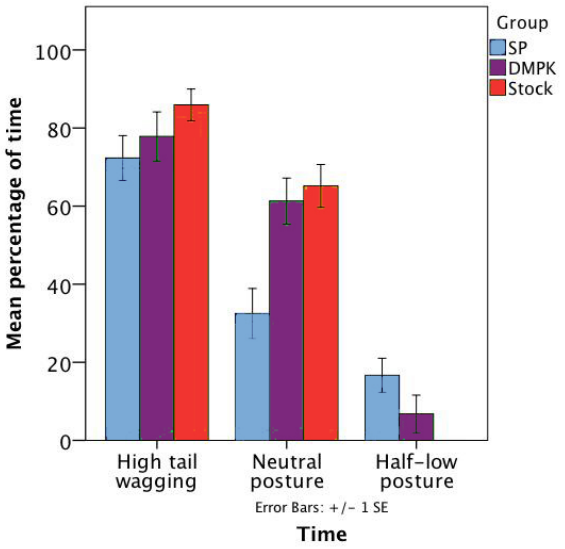


FIGURE 6.9: Mean percentage of time with high tail wagging, high posture and half-low posture during HI by Group

FIGURE 6.10: Mean rate per hour of oral behaviours, paw lifts and total events by Affective State during HI

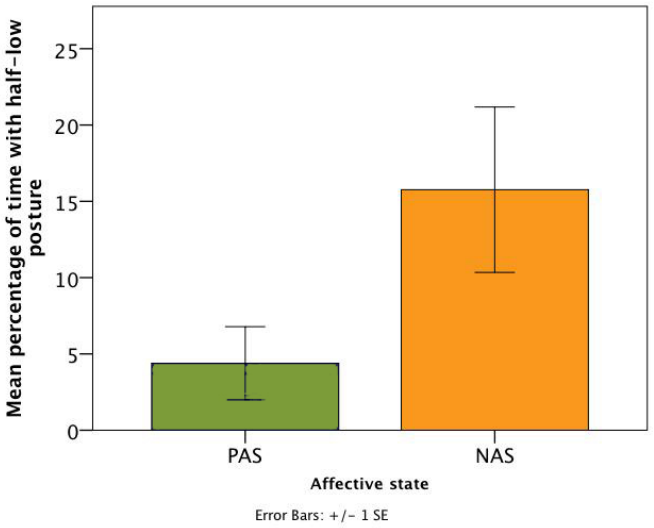


FIGURE 6.11: Mean percentage of time with half-low posture by Affective State during HI

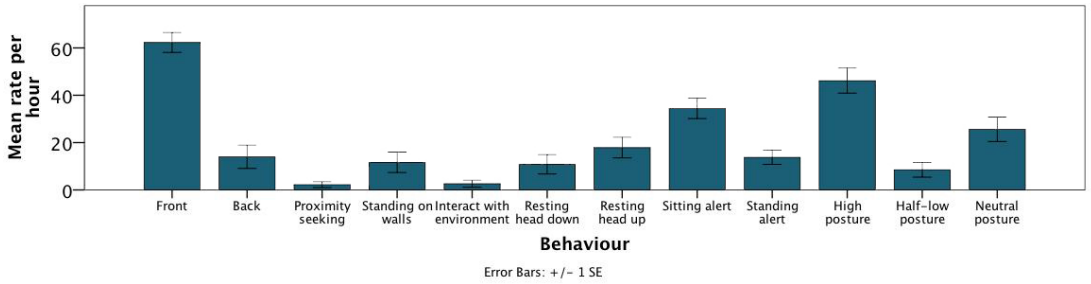


FIGURE 6.12: Mean percentage of time in behaviours during single-housing

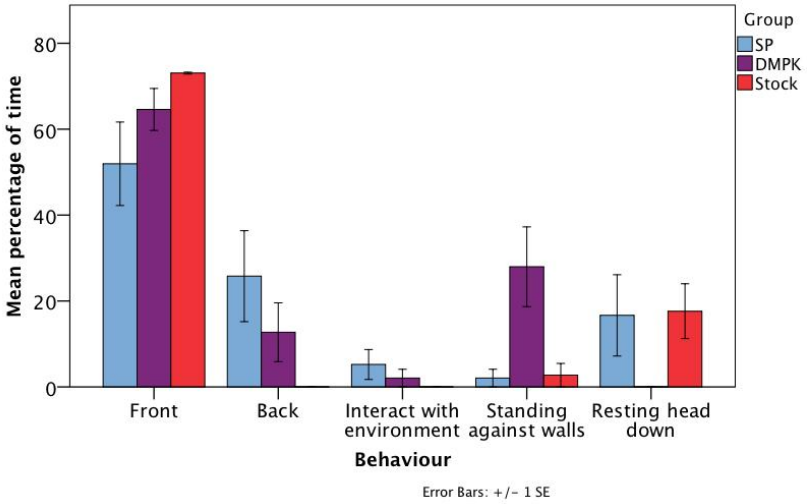


FIGURE 6.13: Mean percentage of time spent in behaviours in home pen by Group during single-housing

6.3.2.2 Single-housing

Dogs spent 62% of time at the front and only 14% at the rear of the pen. Proximity seeking was uncommon, at 2% of time. Interacting with the environment accounted for less than 3% of time and resting behaviours for 28%. Dogs exhibited alert behaviours for almost 50% of time, along with high posture 46%. Neutral posture was exhibited for 25% of time and half-low posture for 8.5%. behavioural events occurred at a rate of 13 per hour, noticeably different from behaviour at baseline.

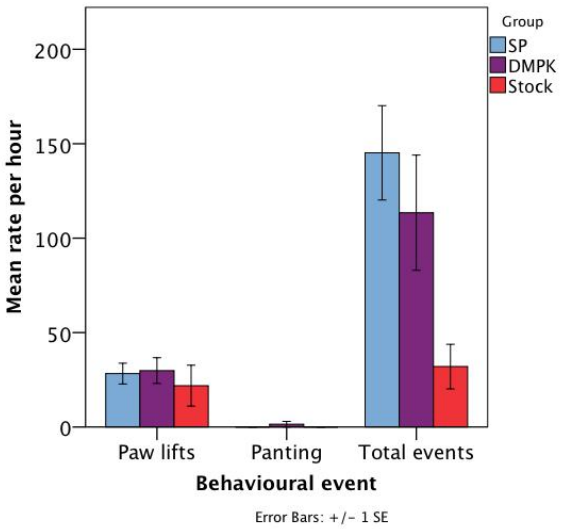


FIGURE 6.14: Mean rate per hour of paw lifts, panting and total events by Group during single-housing

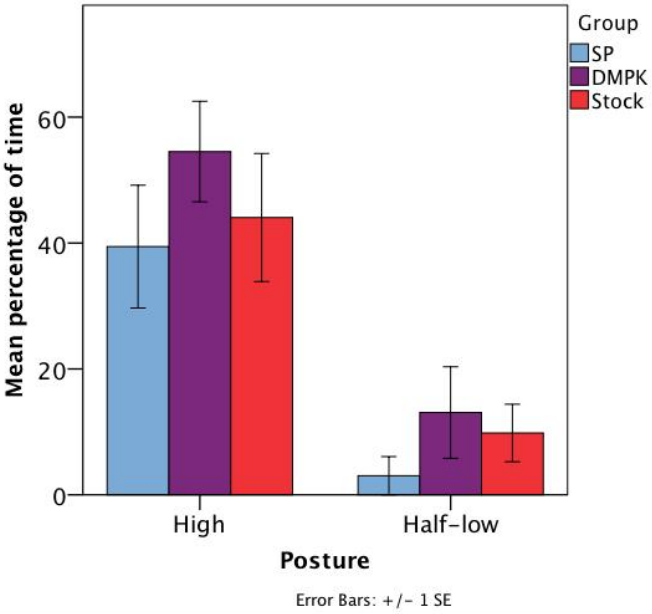


FIGURE 6.15: Mean percentage of time spent with neutral and half-low posture by Group during single-housing

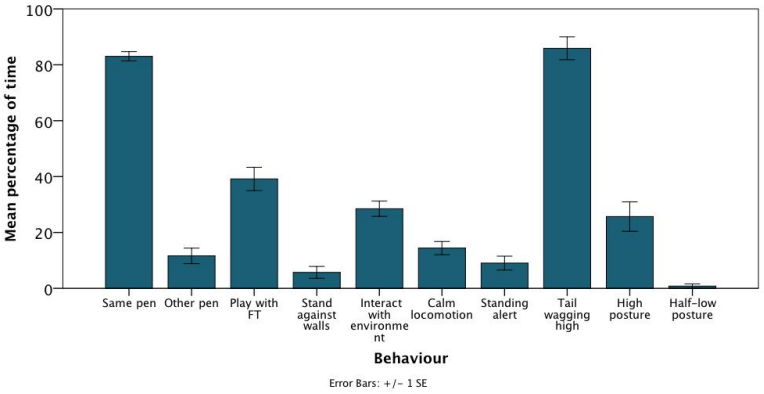


FIGURE 6.16: Behaviour during Feeding Toy

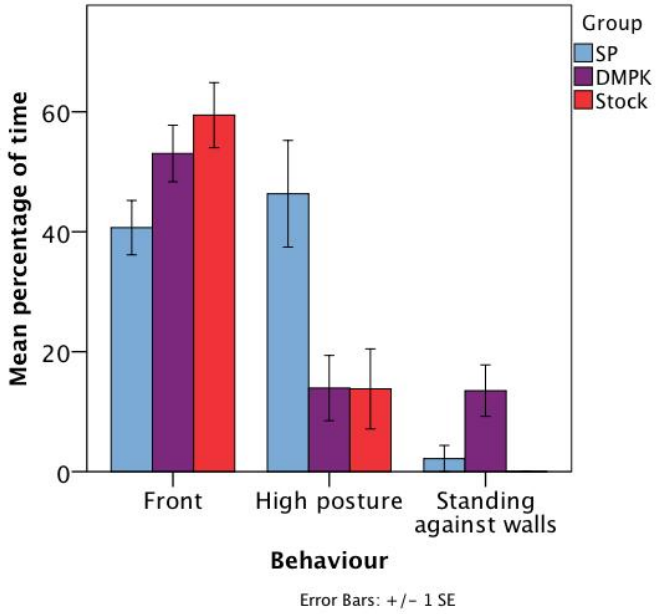


FIGURE 6.17: Mean percentage of time at front of pen, with high posture and standing against walls by Group during FT

6.3.2.3 Feeding Toy

Dogs spent 80.3% of time in the same pen as the FT and 39% playing with it, with 28% of time interacting with the environment. Little time was spent in undesirable behaviours such as standing against walls (5.7%) or standing alert (9%). High tail wagging was exhibited for 86% of time and neutral posture for 56.5%. The mean rate per hour of behavioural events was 5.5. Dogs spent more time in the same pen as the FT than the adjoining pen ($t(22)=17.036$, $p<.001$), suggesting that although it was unfamiliar, dogs did not find its novelty aversive.

There was an effect of AS on interacting with the environment, with NAS dogs

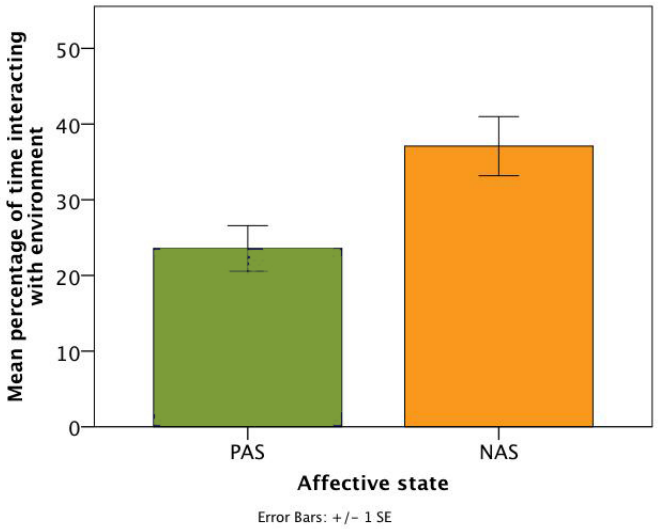


FIGURE 6.18: Mean percentage of time interacting with environment by Affective State during FT

spending more time interacting with the environment (t(20)=2.723, p=.013). This may be explained by a NS trend towards PAS dogs spending more time playing with the FT (p=.068), whilst the NAS dogs were collecting dropped food. NAS dogs may have been less willing to play with the FT as it was unfamiliar.

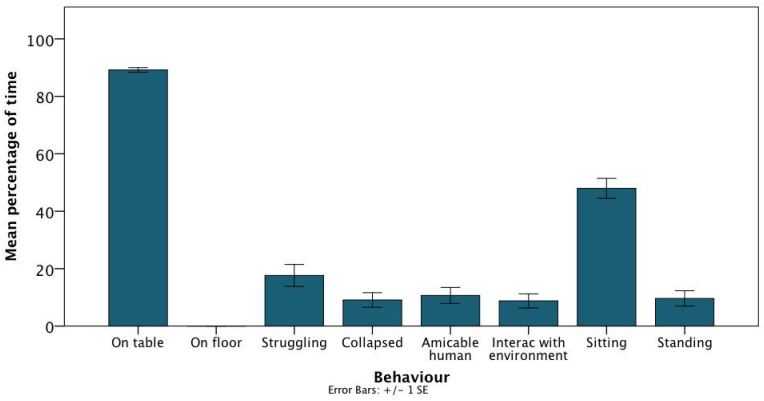


FIGURE 6.19: Behaviour during Restraint

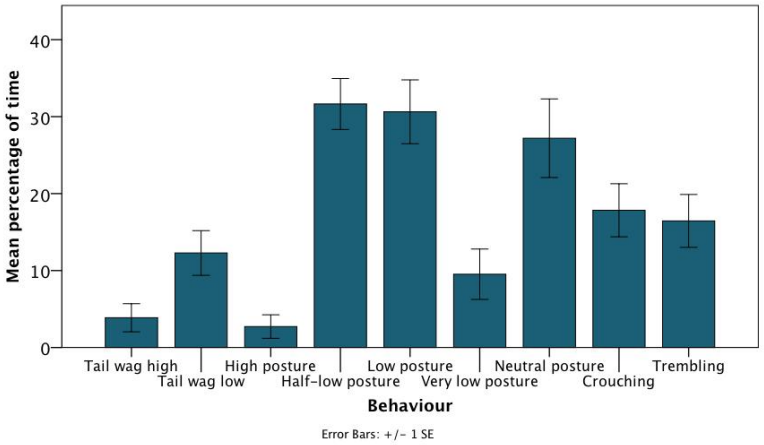


FIGURE 6.20: Posture during Restraint

6.3.2.4 Restraint

Dogs spent most of the restraint period on the table. Dogs spent a considerable amount of time either struggling (17.6%) or collapsed (9%). Dogs spent less than 50% of time sitting (the position in which they were to be restrained). Ten percent of time was spent performing amicable behaviours directed at the experimenter, and 8% of time was spent exploring the environment from the table.

High tail wagging was only present for 3.8% of time, while low tail wagging was present for 12% of time. Dogs only exhibited high posture for 2.7% of time and neutral posture for 27% of time, while half-low posture was exhibited for 31% of time, low posture for 30% and very low for 9.5%. Crouching (17.8%) and trembling (16%) were also present. None of these low postures, crouching or trembling were seen in the home pen. Total behavioural events occurred at a mean of 102.8 per hour, with paw lifts and

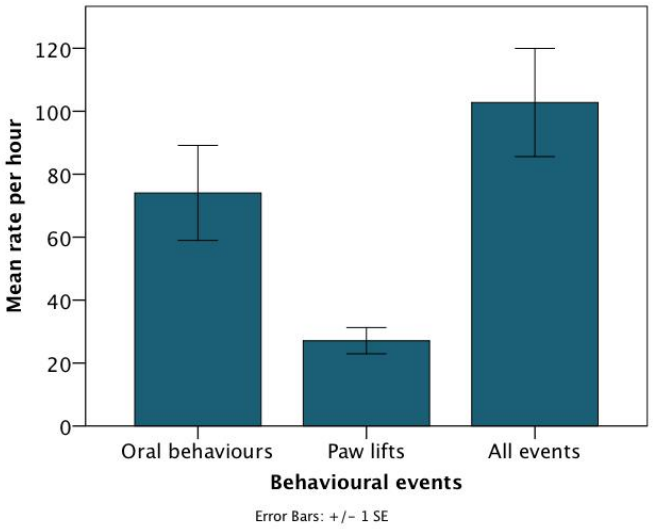


FIGURE 6.21: Behavioural events during Restraint

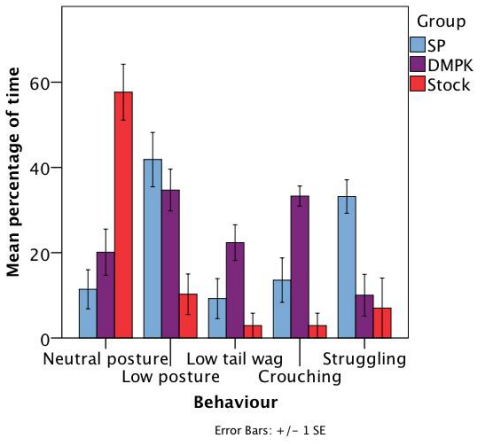


FIGURE 6.22: Behaviour during restraint by Group during Restraint

oral behaviours (both appeasement behaviours) occurring most frequently, at 27 and 74 events per hour respectively.

6.3.3 Cardiovascular response to challenges (n=7)

The principle of harmonising welfare and quality of data output is built upon the concept that increasing welfare, and therefore increasing ability to cope, reduces changes in data output which are caused by factors other than the ones under investigation; welfare is intrinsically linked to the ability to maintain homeostasis in the face of challenge. On this basis, differences in responses to the four challenges would be expected between PAS and NAS dogs; dogs with a greater ability to cope (PAS) are expected to show lower within-subjects and within-group variation in

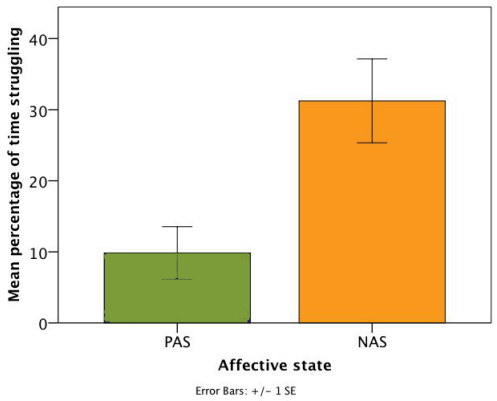


FIGURE 6.23: Mean percentage of time struggling by Affective State during Restraint

response to the challenges, demonstrated through lower variance and lack of changes over time; dogs with a reduced ability to cope (NAS) are expected to show greater within-subjects and within-groups variation in response to the challenges, demonstrated through greater variance and detectable changes over time. The magnitude of between-groups differences are presented below, as are 95% confidence intervals for each of PAS and NAS groups.

6.3.3.1 Human Interaction

There was an effect of time on HR (p=.001) and DBP (p=.002) and a NS effect on SBP (p=.073). There were no differences from Before to After2; nor any interaction between time and AS. From Figure 6.24, it can be seen that the increases in DBP and HR follow the same pattern across time in both NAS and PAS dogs, suggesting that the HI challenge had a similar effect on both groups, although mean HR is slightly higher across all time points in PAS dogs. Across all three parameters, 95% CIs largely overlap, which means that the two affective states respond in a similar manner.

6.3.3.2 Single-housing

There was no effect of time on HR, SBP or DBP (.603<p<.718), and there was no interaction between time and AS (.451<p<.951). Figure 6.25 shows that mean blood pressure values were greater for NAS than PAS dogs throughout the challenge, reflecting the already-detected difference at baseline.

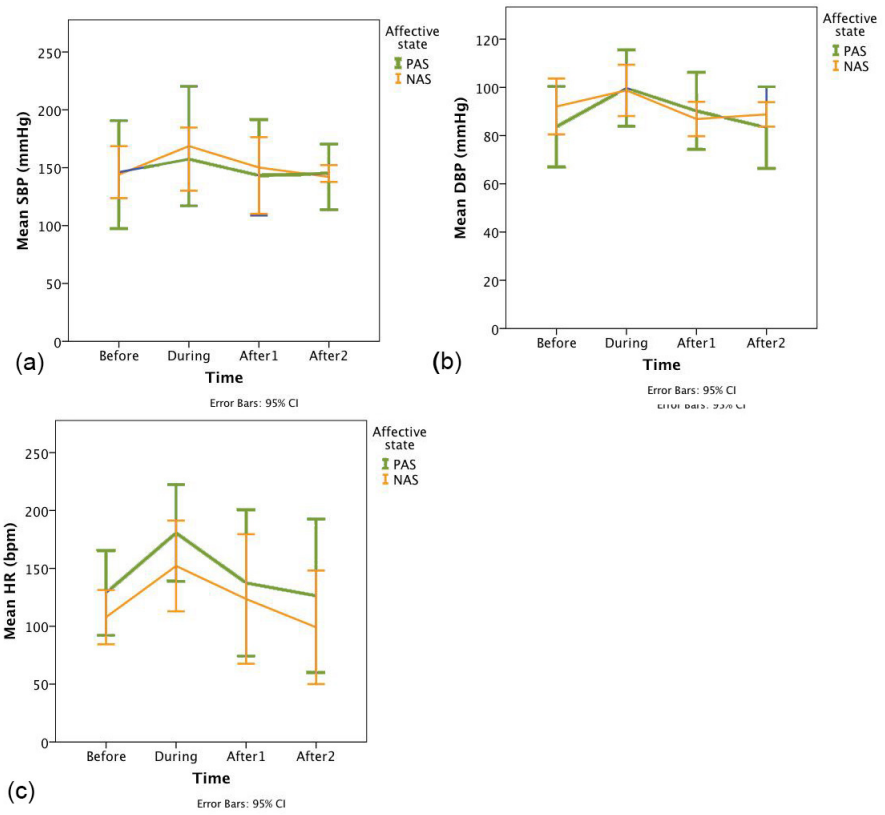


FIGURE 6.24: Changes in (a) SBP, (b) DBP and (c) HR over time for HI Challenge, showing Affective State and 95% CIs

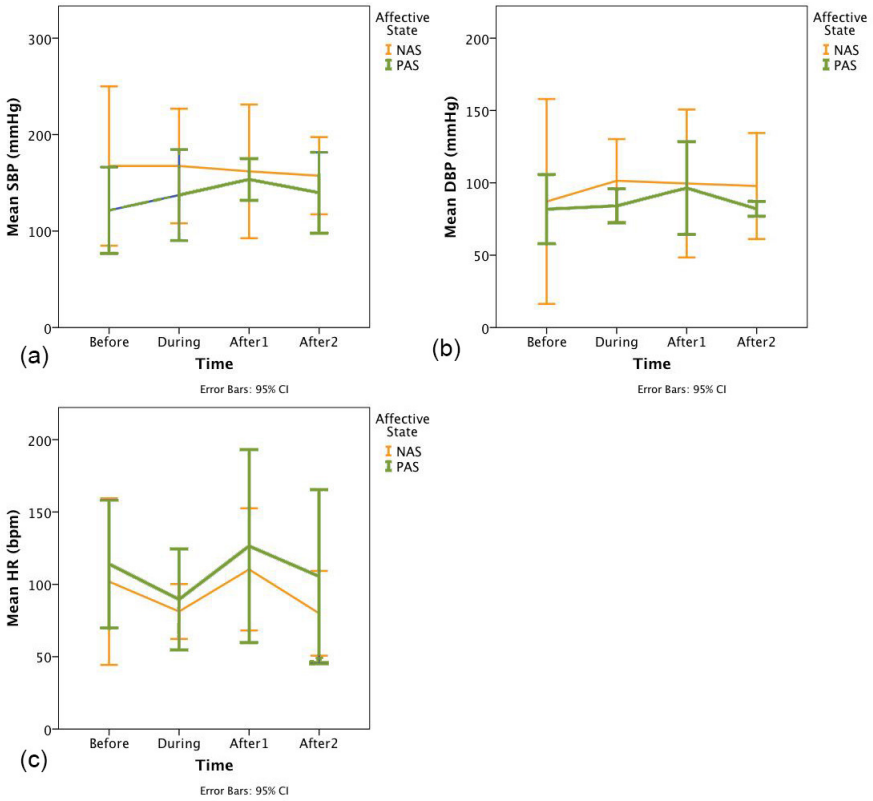


FIGURE 6.25: Changes in (a) SBP, (b) DBP and (c) HR over time for single-housing Challenge, showing Affective State and 95% CIs

6.3.3.3 Feeding Toy

There was no effect of time on SBP or DBP (.840<p<.973). There was a marginally significant effect of time on HR (p=.048) with an increase from Before to During (t(6)=2.439, p=.031) and a decrease from During to After2 (t(5)=2.558, p=.025). There was no difference in HR from Before to After1 or After2. There was also no interaction between time and AS on any parameter (.877<p<.984). Increased HR may be the result of the increase in activity (interacting with the environment, play and time at the barrier). From Figure 6.26, it can be seen that mean HR was higher in the PAS dogs during the FT challenge, possibly due to them interacting more with the FT. The size of the effect of affective state on HR during the challenge was d=1.27, however this returned to a small effect (d=0.47) immediately afterwards.

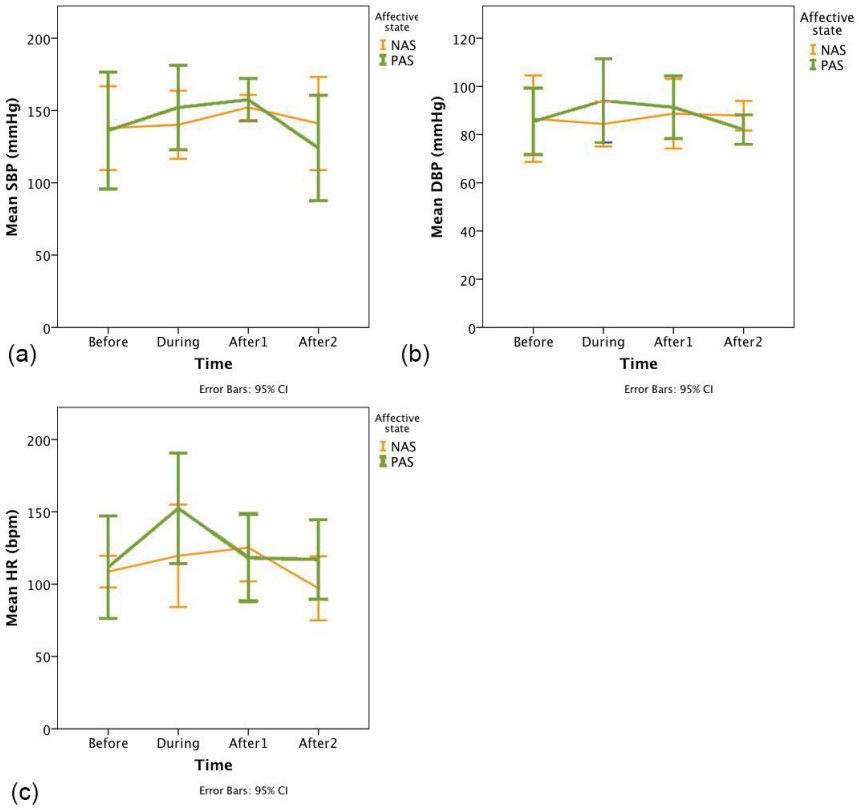


FIGURE 6.26: Changes in (a) SBP, (b) DBP and (c) HR over time for FT Challenge, showing Affective State and 95% CIs

6.3.3.4 Restraint

There was a significant effect of time on both SBP (p=.001) and DBP (p<.001). From Figure 6.27, it can be seen that there was a greater increase in blood pressure during

restraint for NAS dogs than for PAS dogs. There was a significant *effect* of affective state on SBP (p=.002) as blood pressure remained higher throughout, and a significant *interaction* between time and affective state for DBP (p=0.48), with DBP being higher for NAS dogs during restraint. The effect of affective state on SBP and DBP during restraint was considerable, d=2.189 and d=1.627 respectively. Although DBP had returned to baseline by After2, SBP remained raised in NAS dogs, with the effect of affective state being d=1.58 30 minutes after the restraint challenge.

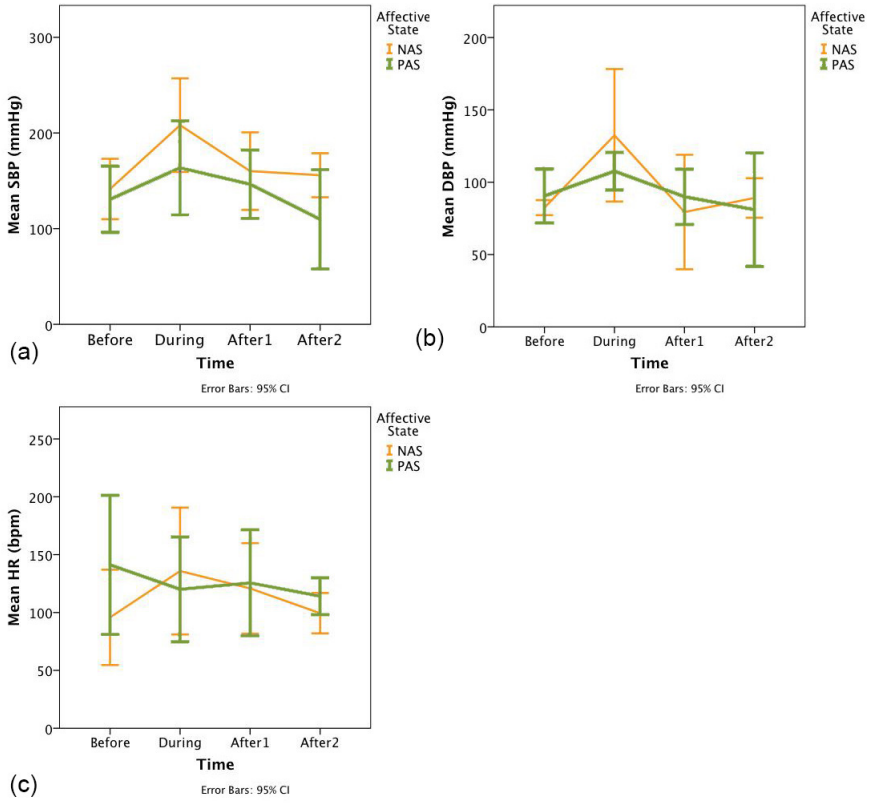


FIGURE 6.27: Changes in (a) SBP, (b) DBP and (c) HR over time for Restraint Challenge, showing Affective State and 95% CIs

6.4 Discussion

6.4.1 Human Interaction

Human interaction (HI) caused significant changes in behaviour for the duration of the 10-minute HI period (posture, resting, tail wagging), but few of these persisted afterwards. While immediate changes in behaviour suggest that HI was a positive experience for all groups, the response to HI was greatest in the dogs which

experienced least human contact, as seen in greater behavioural response (and accompanying CV response) in SP dogs. These dogs experienced much greater arousal in response to the presence of the experimenter (more standing alert, more high and half-low posture) and the level of excitement was so great that it also had an effect on HR and BP. While this level of excitement is undesirable as it introduces unwanted variation, human contact is a necessary part of husbandry and regulated procedures. These data suggest that in order to reduce unwanted increases in HR and BP when staff are present or interacting with dogs, more regular positive interactions with staff would be beneficial, as often suggested in laboratory animal manuals and guidelines (e.g. Prescott et al., 2004). This is most true of SP dogs and NAS dogs, who exhibited more half-low posture and behavioural events in the presence of the experimenter, although this did not translate into behavioural changes before or after HI. The unwanted changes resulting from human contact are also seen in response to restraint, suggesting a common response to human contact, rather than one which is context specific and so strategy to mitigate these effects would benefit from addressing all areas of human contact, rather than positive or negative contact alone.

6.4.2 Single-housing

There were clear behavioural indicators of reduced welfare during single-housing (increase vigilance, decreased restfulness, changes in posture), and while some of these were present at After1, behaviour had returned to baseline by After2. DMPK dogs had a greater response to single-housing, possibly because of the increased association between single-housing and aversive events compared to the other groups; although a brief single-housing appears to be an aversive event, there were no lasting effects. There were no effects of AS or group on behaviour while group-housed, although there was a greater response during single-housing by DMPK dogs. This suggests that brief single-housing presents a similar experience for dogs regardless of optimism or existing welfare, and may be more influenced by the association between single-housing and other events. EU Directive 2010/63 classes single-housing of over four hours as a "substantial severity" for the dog, however it remains unclear at what time single-housing becomes distressing for the dog. From these data, the factors involved in an individual's response to single-housing are related to its history.

Stock dogs did not appear to find single-housing stressful. SP dogs either did not find it stressful, or were so habituated to single-housing and already exhibited more indicators of negative welfare that the effects of single-housing were not detectable. DMPK dogs found single-housing more aversive than the other groups, suggesting that their experience of single-housing influenced their response to it in an ambiguous

context (no clear signals for feeding or regulated procedures). This suggests that in order to Refine the experience of single-housing, the opportunity should be taken at the Stock stage to increase the resilience of dogs to the effects single-housing and desensitise them. The literature also suggests that a predictable signal is preferred for aversive events (Badia et al., 1973). Although negative anticipation is experienced before the aversive events, the increased predictability may increase the ability of dogs to cope with single-housing where it can precede various events.

6.4.3 Feeding Toy

There was a high level of engagement with the FT by all dogs, and unlike HI the effects on behaviour persisted following the removal of the FT (increased resting head up and neutral posture, decreased behavioural events and alert behaviour). Although staff expressed concerns regarding resource guarding on the introduction of a 'valuable' item such as the FT, there were only two instances of competitive behaviour (involving the same two dogs). Although there was a trend towards NAS dogs playing less with the FT than PAS dogs, and an increase in HR was seen, it is possible that more frequent contact with the FT would reduce the novelty of it and allow NAS dogs to interact with the FT more. Unlike HI, the positive effects of the FT persisted for the duration of the experiment which suggests that the FT may be a simple and effective way of improving welfare.

6.4.4 Restraint

Of the four challenges, restraint had the greatest effect on behaviour and cardiovascular parameters. Restraint had a negative impact on dogs for the duration of the restraint, although less so for Stock dogs who did not exhibit the low posture and increase in behavioural events seen in other groups. The differences in past associations with restraint for the three groups is likely to have led to this; Stock dogs would have been lightly restrained for health checks but not for regulated procedures.

Restraint also increased the level of anxiety in the home pen (decrease in resting, and amicable behaviour). NAS dogs had a greater response to restraint both during and after restraint, with more time struggling and more behavioural events on the table and more alert behaviour in the home pen. Although all dogs had a CV response, NAS dogs had a greater cardiovascular response to restraint which lasted for up to 30 minutes afterwards. The magnitude of the effect of affective state during and after restraint was considerable, and most importantly was still evident in SBP 30 minutes after restraint. Recordings of telemetered data may not typically be taken immediately

after dosing, but a change which persists beyond 30 minutes may influence data collected.

The maximum values obtained for SBP, DBP and HR are close to the maximum values suggested by Van Citters and Franklin (1969) as the maximum values possible for highly fit dogs under extreme physical stress. Unlike the dogs in that study, these dogs were not extremely fit, and nor were they subject to exertion under physical exercise. It has been shown that sudden increases in blood pressure and heart rate (for example, those experienced in an acute response to psychological stress) are more damaging than those occurring as the result of a highly-fit individual exercising, even when the maximum values achieved are the same. What is most concerning about this response is that the dogs did not undergo any regulated procedure, nor did they have any signal that one might be about to occur. I had never been present during a regulated procedure, and there was no equipment present which might suggest that one was about to occur. In addition, the dogs had only previously encountered me in a positive setting - either presenting them with human contact opportunities or food treats. It seems reasonable to assume that if such a response were seen in the context of a non-regulated and not potentially distressing procedure such as gentle restraint, a much greater response would be seen in response to a painful or aversive experience. Regular cardiac stress to this extent may affect physical health, other organ systems, the quality of data obtained from the dogs and the cumulative experience of the dogs (see Chapter 2). Although there is no post-mortem evidence of this occurring reported in this population, restraint appears to have the ability to compromise data quality.

The between-groups differences, with the group experiencing most frequent regulated procedures (SP) having a greater response and the group which had never experienced regulated procedures (Stock) showing little response, points to the experience of regulated procedures and husbandry factors having an influence on the effects of restraint. Although there was no association between the experimenter and regulated procedures, the anticipation of regulated procedures caused by restraint elicited a strongly negative response from some dogs.

6.4.5 A Welfare Assessment Framework

The data gathered in Chapters 4-6 have shown a consistent pattern of behaviour, cardiovascular function, affective state and mechanical pressure threshold, demonstrating two distinct welfare types. The aim of developing a Welfare Assessment Framework was to provide a means to identify undesirable patterns of data output (such as the reduced sensitivity and repeatability seen in cardiovascular data) using

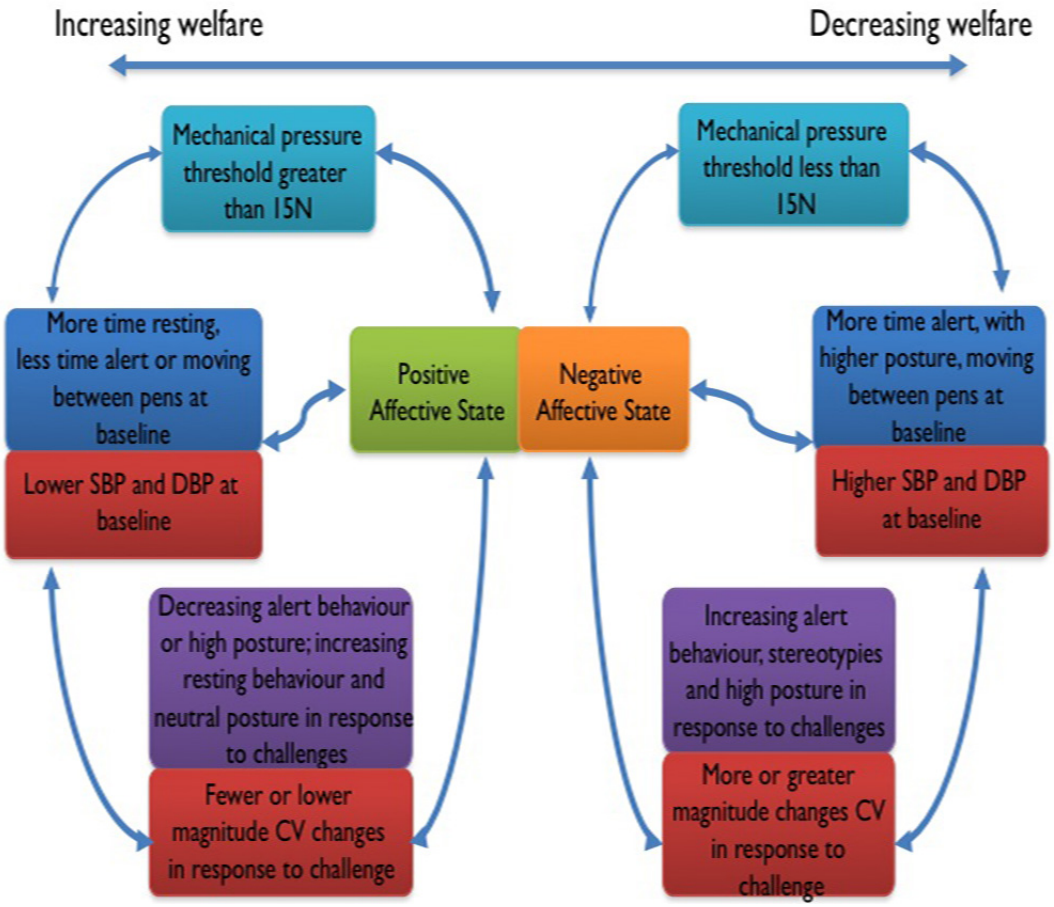


FIGURE 6.28: The Welfare Assessment Framework depicted diagrammatically

behaviours easily identifiable in the home pen. Using Affective State as a central measure, Figure 6.28 depicts the relationship between each of the elements which makes up the Framework.

Positive welfare is associated with a positive Affective State, MPT >15N (when using a 10mm tip), more restful behaviours in the home pen with neutral posture, lesser and more positive reaction to behavioural challenges, lower blood pressure in the home pen and less cardiovascular reactivity in response to challenges. Conversely, poor welfare is associated with a negative Affective state, MPT <15N (when using a 10mm tip), more vigilant or restless behaviour in the home pen associated with more high posture, a greater and more negative response to challenges, blood pressure is higher in the home pen and there is greater cardiovascular reactivity to challenges.

Taken together, the patterns found in dogs with positively- (PAS) and negatively-valenced (NAS) welfare suggest that there should be concern regarding the data output obtained from their use. Dogs which are more sensitive to physical sensation or pain (MPT) are likely to suffer reduced welfare as a result of regular

handling and regulated procedures. NAS dogs which have a pattern of vigilant, more active behaviour in the home pen may have different physiological responses such as immunity or endocrine function (see Chapter 2). The increase in blood pressure and heart rate demonstrated by some dogs in response to brief restraint is capable of causing target organ damage (Brown et al., 2006), with acute responses being more damaging to physiology than gradual adaptations. While no patterns of organ pathologies relating to blood pressure were reported in the colony from which these dogs were drawn, subtle organ pathologies may influence the confidence with which changes are attributed to a test compound, influencing the regulatory outcomes. What is clear is that mild regulated procedures have the capacity to cause changes in welfare and physiology. Where these procedures are combined with other factors such as housing, husbandry protocols and more severe procedures, these short-term changes in physical health may confound results.

Without the ability to identify the dogs at risk of exhibiting poor welfare and either removing them from a study or implementing additional Refinements to improve welfare, this could have a serious impact on quality of data output, and for quality of science practice as unwanted variation is introduced and dog use is likely to be higher as a result. These data support both the relationship between welfare and data quality, and the ability of Refinements to increase welfare to mitigate unwanted changes.

Appendix E shows a Welfare Monitoring Tool; using the behaviours most indicative of positive and negative welfare, the tool is weighted to provide scores which reflect welfare in the home pen using data from Chapters 5 and 6. This tool is implemented in the Chapter 7 to investigate the ability of the Framework to detect the effects of Refinements.

CHAPTER 7

Refining oral gavage in the dog

"It is widely recognised that the 'humanest' possible treatment of experimental animals, far from being an obstacle, is actually a prerequisite for successful animal experiment"

W.M.S Russell and R.L. Burch
(1959)

Abstract

Oral gavage is a technique frequently used to deliver a compound directly into the stomach. As with other animals, in the dog, gavage is aversive and the frequency of its use is a cause for welfare concern but little research has been published on the technique nor how to Refine it. Using the Framework, this study compares the effects of sham dosing (used to attempt to habituate dogs to dosing) and a Refined training protocol against a control, no-training group to determine the benefit to welfare and scientific output of each technique. The pattern of findings show that sham dosing is ineffective as a habituation technique and 'primes' rather than desensitises dogs to dosing. Dogs in the control group show few changes in parameters across the duration of the study, with some undesirable changes during dosing, while dogs in the Refined treatment group show improvements in many parameters. It is recommended that if there is no time allocated for pre-study

training a no-sham dosing protocol is used. However, brief training periods demonstrate a considerable benefit for welfare and quality of data to be obtained from the dogs' use.

7.1 Introduction

7.1.1 Oral gavage as a dosing technique

Oral gavage is a technique for delivering a substance directly into the stomach and is frequently used to administer novel compounds in research and toxicology testing, as the oral route is the most common route in humans. It is recognised as an invasive and aversive event in the life of a laboratory animal (Wallace, Sanford, Smith & Spencer, 1990). In a standard one- or three-month toxicology study, dogs may experience up to daily oral gavage, while pharmacokinetic studies may use more than one dose in a day. Whilst it is recommended that dogs are introduced to the technique and habituated (commonly referred to as Sham Dosing, SD) before a study begins, there is little standardisation in the method for doing this.

In addition there is no robust scientific evidence that demonstrates any welfare benefit from the procedure of SD, apparent cooperation may be a 'freezing' response to fear. A proficient technician is able to deliver the dosage of a compound quickly and without physical trauma. However, a technique which is invasive and which happens at potentially unpredictable intervals and outwith the control of the dog always has the potential to be highly aversive (Laule, 2010).

There is comparatively little guidance published on training of the laboratory-housed dog for procedures (i.e. organisations such as NC3Rs and IAT produce guidance for procedures in rodents) and almost nothing specifically for the Refinement of oral gavage in the dog. However, there is a wealth of literature available (e.g. McKinley, Buchanan-Smith, Bassett & Morris, 2003; Prescott & Buchanan-Smith, 2007; Laule, 2010) supporting the benefits of positive reinforcement training (PRT) for various aspects of husbandry and procedures for many species in the laboratory environment. PRT is also used extensively in the training of dogs in other situations (e.g. Fjellanger, Andersen & McLean, 2002; Hiby, Rooney & Bradshaw, 2004; Batt, Batt, Baguley & McGreevy, 2008a, sniffer, pet and guide dogs respectively).

Our previous research using other groups of dogs in the same facility identified convergent validity in patterns of behaviours, cardiovascular parameters, affective state and mechanical pressure threshold which identified welfare differences between dogs (see Chapters 4-6). Those with more negative welfare showed higher levels of

undesirable behaviours (and often more 'reactive' behaviours) at baseline in the home pen and in response to behavioural challenges; higher blood pressure at baseline and a greater cardiovascular response to a brief restraint, a negative affective state (NAS) and a lower threshold for tolerance of a mechanical pressure stimulus. It is likely that these dogs adapt less well to an aversive technique such as gavage. Anecdotally, technical staff report an unidentified factor which causes some dogs to fail to adapt, which is likely to produce unwanted variation and poorer quality data output. There is also a concern that stress-induced vomiting, reported to be as high as high as 50% (Gad, 2006), can affect experimental outcomes. Understanding the link between quality of data and other indicators provides a means of monitoring the impact of changing welfare. The Framework is designed to identify those dogs most at risk of negative welfare and highlights the need for *harmonisation* of training and desensitisation, by providing sufficient Refinements to allow dogs to adapt to and cope with regulated procedures.

Quality of scientific output in cardiovascular data was also found to be lower (increased variability, lower repeatability, see Chapters 5 and 6) in NAS dogs. As a pattern of behaviours in these NAS dogs was identified and also seen in other groups of dogs with a NAS, it may be possible that other dogs exhibiting this pattern of behaviour also have lower quality of scientific output. The response to a restraint as part of the development of this Framework highlighted it as an aspect of study protocol particularly in need of Refinement, given the undesirable change in behavioural and cardiovascular parameters seen in the absence of a regulated procedure.

7.1.2 Training for procedures through habituation and desensitisation

7.1.2.1 The distinction between habituation and desensitisation

Habituation is the process by which the response to a stimulus diminishes by repeated exposure to the stimulus, while desensitisation is the process of reducing the response to an aversive stimulus by pairing a reward with the presentation of the stimulus (Laule, 2010). While habituation may be common practice for regulated procedures in a laboratory setting and results in a decreased behavioural response to the aversive stimulus or event, this may not represent actual habituation but rather a "freezing" response and cooperation, while internally arousal has not decreased (see Ruys, Mendoza, Capitanio & Mason, 2004). It is commonly recommended that some form of "habituation" take place before a study (e.g. Laule, 2010), however the interpretation of its use varies and there is currently no standardisation in the use of desensitisation within the laboratory environment for the dog (Prescott et al., 2004). This may be due

to lack of understanding of the distinction between habituation and desensitisation, and lack of a structured programme to implement desensitisation. One commonly-given reason for this is the perceived introduction of variability in data due to interactions between non-standardised food in the form of treats during *in vivo* testing, and as such negative reinforcement (NRT) training is often more commonly used than positive reinforcement. NRT may involve the use of an unpleasant stimulus and as such encourages fear, resistance and avoidance, all of which are undesirable states in an *in vivo* model of a healthy human.

7.1.2.2 Control and predictability, and the influence on welfare

In an environment where unpleasant stimuli are unavoidable, control and predictability are especially important to animals. Control is the ability to make a decision which changes the response of something in the environment. Weinberg and Levine (1980) defined control as "the ability to make active responses during aversive stimulus", while Sambrook and Buchanan-Smith (1997) defined it as "the difference in likelihood of an event occurring depending on an animals' behaviour". Overmier, Patterson and Wielkiewicz (1980) stated that this ability to exert control increases the positive effects and decreases the negative effects of an event. Therefore control may reduce the negative effects of an aversive event. Control and predictability are also interlinked, as increased control leads to increased predictability over the occurrence of an event, while increased predictability can lead to an increased ability to exert control, although some aversive events may never be controllable.

There is considerable literature supporting the value of control and predictability for animals' welfare, as discussed in a review by Bassett and Buchanan-Smith (2007), who also found that predictability of a signal is more important to welfare than its temporal relationship, highlighting that that to effective signal must be highly predictive of an event's occurrence to be of benefit. In the laboratory environment many signals are unreliable and therefore not predictable, such as the appearance of staff (signalling husbandry or regulated procedures) or transfer to single housing (signalling feeding or imminent dosing).

Positive reinforcement training (PRT) is a commonly used technique for desensitising laboratory animals while also increasing control by giving the animal the choice to cooperate rather than by force or coercion, as is the case in the use of negative reinforcement training (NRT) or punishment (Laule, 2010). As PRT is likely to have a more positive impact on welfare than NRT, and is also likely to increase rather than

decrease cooperation, it should be the preferred training method in the laboratory environment.

7.1.3 Aims

The overall aim of this study was to determine if the use of sham dosing as a habituation technique has a welfare benefit over no pre-study habituation, and to measure the potential benefits to welfare and quality of science implementing a Refined training and sham dosing protocol. Dogs from three treatment groups were compared (see Table 7.1).

The first goal of this study was to compare the current sham dosing procedure (SD group) with a group receiving no sham dosing (Control group) to determine if the sham dosing procedure alone has a welfare benefit to dogs' welfare. The second goal was to compare both of these groups with a third group receiving Refined desensitisation and handling (RP group) to determine if additional training and Refinements to the sham dosing technique have any benefit to dogs' welfare and quality of scientific output. It was expected that this third (RP) group would demonstrate the highest levels of welfare and corresponding highest quality of data output as measured using welfare-indicative behaviours and stability in body weight, food consumption and heart rate during the Training Phase and during the Dosing Phase.

In addition, the time investment required for each of these groups, as total time to train (Training and SD Phases) against total time to dose (Dosing Phase) was assessed. A staff Welfare Monitoring Tool in the form of a tick sheet was also used and compared to other data obtained with the aim of providing a simple tool which can be used to monitor welfare in future dog studies.

It was anticipated that in SD Phase, dogs in the SD Group would show greater behavioural signs of distress during sham dosing than those in the RP Group. Despite the training given during the Training Phase, as gavage is an aversive procedure, it was unlikely that RP Group would not show an aversive response. It was also anticipated that in the Dosing Phase, dogs in RP Group would show higher levels of positive-welfare indicating behaviours (resting, neutral posture) and lower levels of negative welfare indicating behaviours (vigilance, stereotypies, high/low posture) than those in Control and SD Groups. As a result of a reduced behavioural response to dosing, dogs in RP Group should also show greater stability in measures of body weight and food consumption and a reduced increase in heart rate during dosing than dogs in Control and SD Groups. It was expected that while RP Group would require greater time investment in the Training Phase and SD Phase due to the frequency of

training sessions, time required for dosing in the Dosing Phase would be less than for Control and SD Groups.

7.2 Methodology

The following sections set out the overall design of the study as well as the methodology applied to each group. A detailed study schedule is provided in Appendix F. Details of the dogs in each Group are provided in Chapter 3.

7.2.1 Overview of study design

Table 7.1 illustrates the treatment given to each of the three groups in each aspect of the study.

Table 7.1: Treatment delivered to each of three groups

Group	Control Control	SD Sham dosing	RP Refined protocols
Treatment			
Health check	Once weekly in all Phases	Once weekly in all Phases	Once weekly in all Phases
Training sessions	None	None	4x in Training Phase
Modifications to handling	None	None	All Phases
Predictive signal for dosing	None	None	All Phases
Sham dosing	None	Twice in SD Phase	Twice in SD Phase (Refined technique)
Vehicle dosing	Daily in Dose Phase	Daily in Dose Phase	Daily in Dose Phase

Control Group

Current company practice is to sham dose dogs prior to studies, however some technicians report that sham dosing primes the dogs' response and increases the aversiveness of the event. Control Group received no SD or additional handling outwith the standard weekly health check.

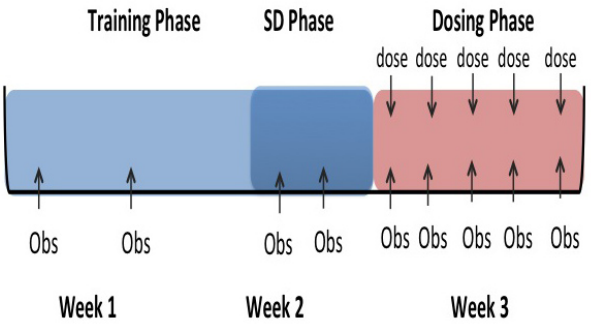


FIGURE 7.1: Schematic depiction of behavioural observations and dosing periods for Control Group

SD Group

SD is recommended to habituate dogs to dosing prior to the beginning of a study and is the current method of habituation used at the Industrial Partner. Dogs are restrained, dosed using a gavage tube dipped in warm water, then returned to the home pen. As this technique is invasive, aversive and does not provide desensitisation or control on the dogs' part, it is anticipated to increase arousal in anticipation of future dosing.

RP Group

Whilst there is disagreement over whether sham dosing or no pre-study habituation is better for welfare in dogs, there is clearly still Refinement needed to several aspects of oral gavage dosing, as evidenced by anecdotal accounts from several facilities highlighting the failure of dogs to adapt to dosing by oral gavage. Based on a review of current literature and in discussion with staff in various roles within Laboratory Animal Sciences, the aspects of PRT, desensitisation and predictability and control described below were identified as the most suitable for use in the laboratory setting, whilst giving the greatest benefit.

7.2.2 Training Phase

Control Group

During the Training Phase dogs were group-housed and were allowed access to an indoor play area during daily husbandry. Once weekly, dogs were be taken to the procedure pod for a physical examination, health check and body weight measurement. No other training or interventions was given during this time.

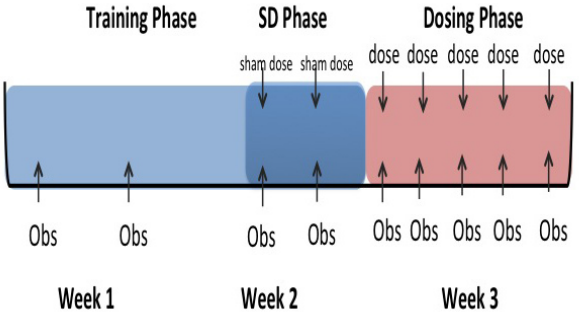


FIGURE 7.2: Schematic depiction of behavioural observations, sham dosing and dosing periods for SD Group

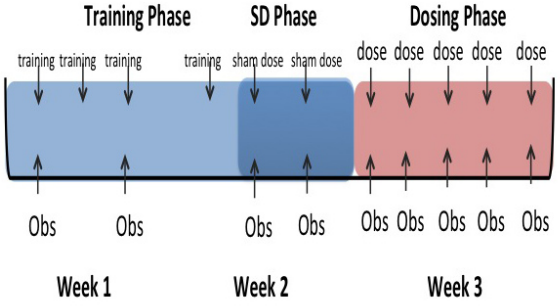


FIGURE 7.3: Schematic depiction of behavioural observations, sham dosing and dosing periods for RP Group

SD Group

Dogs received the same treatment as dogs in Control Group. In addition, dogs were taken to the procedure pod twice in SD Phase, restrained by the handler and dosed by the animal technician using the standard SD technique.

RP Group

During the Training Phase dogs in RP Group underwent three training sessions per week which incorporated predictability and control, PRT and desensitisation. A Refined sham dosing technique was introduced on two occasions in SD Phase at the same intervals as those in SD Group receive sham dosing.

RP Group received several additional Refinements during the Training Phase not administered to Control and SD Groups. These are detailed below.

Predictability

There are currently no reliable signals to indicate imminent dosing. Dogs with a positive affective state (PAS) demonstrate lower levels of arousal in the home pen and in response to behavioural challenges and stimuli in the environment. However, dogs exhibiting a negative affective state (NAS) had higher levels of arousal and more variable heart rate and blood pressure, show persistence of alert behaviour without the ability to return to restful states (see Chapter 6). With no reliable signals for dosing, NAS dogs are at greater risk of spending time in aroused states. In order to provide a reliable signal for dosing, and therefore a shorter period of arousal, a visual signal was provided individually to each dog two minutes before dosing. Previous cognitive bias testing (see Chapter 4) has shown that laboratory-housed beagles quickly learn to differentiate between black and white stimuli, a piece of A4 card was fixed to the front of the pen showing white, and was flipped two minutes before dosing (or training periods) to show black. It was flipped again to white when the dog was returned to the pen following training or dosing. This was designed to provide a brief, controlled period of increased arousal immediately before dosing, decreasing the length of time spent in undesirable behavioural and affective states.

Handling and control

Scruffing and removing a dog from the home pen for dosing removes choice and a sense of control for the dog. To mitigate this (as advised in Prescott et al., 2004), dogs were not scruffed, but were encouraged to approach the handler before lifting and removal from the pen. While the dog ultimately does not have a choice about removal from the pen or dosing, it is the perception of choice which allows a sense of control. Many dogs continued to voluntarily approach the handler throughout the study.

PRT and desensitisation

Dogs received four five-minute training sessions in the Training Phase. During these sessions a programme of desensitisation using PRT was followed and a training schedule is provided in Appendix G. Training progressed from training for calm restraint on the table, to introduction to gavage equipment to dosing. Dogs were therefore to be desensitised to the presence of the two members of staff involved in dosing (handler and technician), being on the table, being restrained, the sights, sounds and smells of equipment and the Refined sham dosing technique. The target behaviour was calm sitting on the table during restraint. Vocal praise, calm touch and a food treat (Pedigree®Cheesy Bites) were used as primary reinforcers and were used both in the procedure pod and immediately upon return to the home pen. Progression through each of the stages of training depended on the responses of the dogs (as some dogs may desensitise to staff and restraint more quickly than others) but the training programme was designed such that all dogs would be exhibiting the target behaviour by the end of the SD Phase. In order to provide a comparable number of sham doses

to SD Group, dogs in RP Group received two sham doses on the same two days as those in SD Group.

7.2.3 SD Phase

7.2.3.1 Control Group

The Control Group received no interventions during the SD Phase. Observations were made at times corresponding to before and after sham doses for the other groups.

7.2.3.2 SD Group

On each of the two days of the SD Phase, the standard sham dosing protocol was followed. Dogs were separated into individual housing shortly before sham dosing began. Dogs were then taken one at a time in a pre-determined order to the procedure pod nearest the pen; once positioned on the table and restrained by the handler, the technician inserted the gavage tube which had previously been dipped in warm water to make it more malleable. No dose was administered. Dogs were immediately returned to the home pen and allowed to return to group housing once the last dogs had received its sham dose.

7.2.3.3 RP Group

Standard company protocols for dosing by oral gavage required the gavage tube to be dipped in warm water before use. In addition, palatable paste (Beaphar®Vitamin Malt Paste) was used to coat the gavage tube before the sham dosing procedure. The aim of this was to encourage swallowing rather than regurgitation when the tube was inserted, reducing the physical discomfort associated with dosing. This paste was only used in the sham doses given in the SD Phase and never in the Dosing Phase. The paste was highly palatable and the residue from the tube also acted as an instantaneous reinforcer following sham dosing. Dogs were rewarded with a food treat following sham dosing.

7.2.4 Dose Phase

On each day of the Dosing Phase, all Groups were dosed with hydroxypropyl methyllcellulose (HPMC) at 2 ml kg^{-1} as per the dosing protocol of a standard

toxicology study. During the Dosing Phase, dogs in Control and SD Groups underwent the identical treatment, while dogs in RP Group also continued to receive Refined handling and predictable signal as detailed above. No additional food treats were administered to RP Group during the Dosing Phase.

7.2.5 Other measures (all Groups)

Table 7.2 shows measures which were administered to all Groups, at the time points described.

Table 7.2: Measures collected for all Groups

Measure	Description	Predicted results
Food consumption (FC)	Daily food consumption of 300g standard dry diet	FC predicted to be more stable in RP Group
Body weight (BW)	Body weight measured once weekly during health checks	BW predicted to be more stable in RP Group
Time investment	Total time taken for training in the Training Phase and total time to dose in the Dosing Phase	Time investment is higher for RP Group in the Training Phase but predicted to be lower during the Dosing Phase
Heart rate	Heart rate (bpm) obtained for all dogs during the Dosing Phase	Heart rate was predicted to be more stable and show lower magnitude increases during dosing for RP Group
Mechanical Pressure Threshold (MPT) Testing	MPT was taken using the protocol outlined in Chapter 4 on Days 11, 12, 15, 17 and 19	MPT was predicted to be more stable in RP Group
Welfare Monitoring Tool (WMT)	A score sheet designed for use by the technician (Appendix E) with the aim of providing a means of monitoring welfare	Scores for RP Group were predicted to be lower (better welfare) and more stable than other Groups across the study
Visual analogue scale (VAS)	A score given using a 0-10cm line on each day of dosing for each dog, representing the ease of dosing, with 0 being the worst and 10 the best score	VAS scores were predicted to be higher (easier to dose) for RP Group than for other Groups

7.2.6 Data analysis

7.2.6.1 Behavioural observations

All behavioural observations conducted in the home pen were of five-minute lengths. Observations were taken before and after training, sham dosing and dosing sessions. This allowed comparisons between- and within-Groups to be made across the three Phases (Training, SD and Dosing) and also before and after sessions. Behavioural observations were also conducted during all sessions. All behavioural observations were recorded on a camcorder and behaviours were scored remotely using The Observer 10.5XT. Home pen behaviour was scored using a combination of instantaneous (30 second intervals) and continuous sampling and behavioural states are presented as a percentage of time, while behavioural events are presented as a rate per hour. All other behaviours were recorded using continuous sampling only due to short durations. Behavioural states are presented as a percentage of time and behavioural events as a rate per hour. Data which were not normally distributed were transformed using an angular transformation.

Inter-rater reliability was conducted by a rater blind to group allocation to ensure the validity of measurements. Two video samples of five-minute duration from each Group were used, with 30 second intervals for instantaneous sampling, of behavioural states and continuous sampling of behavioural events. The same coding scheme was used for both raters. The proportion of behaviours which had been scored in agreement was calculated using the inter-rater reliability function of The Observer XT and was found to .8. The samples for which there was disagreement were found to be posture, which the second rater found difficult to identify.

All data were entered into SPSS 19.0 for Windows and analysed using the methodologies laid out in Chapter 3.

7.2.6.2 Welfare Monitoring Tool

There were two scores available for each day - a home pen score and a dosing score. The maximum score possible each hour was 40, and therefore 280 for each full day. Raw daily scores were entered into a spreadsheet for analysis for analysis as above.

7.2.6.3 Visual analogue scales

On each day of the Dosing Phase, the technician marked on a 10cm line a score representing the ease of dosing for that dog. These scores were measured and raw data (0-10) entered into a spreadsheet for analysis as above.

7.2.7 Ethical approval

This study was conducted in compliance with A(SP)A (1986), with all relevant licenses being held by staff at the Industrial Partner. Ethical approval was sought from the Ethical Review Committee before commencing with the study, and further ethical approval was then granted by the Psychology Division's Ethics Panel.

7.3 Results and interpretation

7.3.1 Between-Groups differences in home pen behaviour during the Training Phase

ANOVAs were conducted with one between-subjects factor of Group (3 levels) and one within-subjects factor of Day (4 levels), followed by planned post-hoc t-tests.
Non-normally distributed data were analysed using a Kruskal-Wallis test followed by Mann-Whitney U tests.

7.3.1.1 Behavioural states

During the Training Phase, there was an effect of Group on the following behaviours:

Table 7.3: Results of ANOVAs between Groups during the Training Phase

Behaviour	F(2, 93)	p	Main findings
Resting head down	7.765	.014	RP>C,SD
Sit alert	5.762	.004	C,SD>RP
High posture	4.434	.015	SD>C,RP
Neutral posture	5.753	.004	C,RP>SD

Where significant effects of Group were found, planned post-hoc t-tests and Mann-Whitney U tests were conducted to determine where between-Groups differences lay. The results are shown in Tables 7.10 and 7.8.

TABLE 7.4: Results of Kruskal-Wallis tests between Groups during the Training Phase

Behaviour	χ^2**(2)**	**p**	**Main findings**
Back	6.857	.032	RP>C,SD
Tail wagging high	10.848	.004	RP>C,SD
Half-low posture	8.801	.012	C>RP

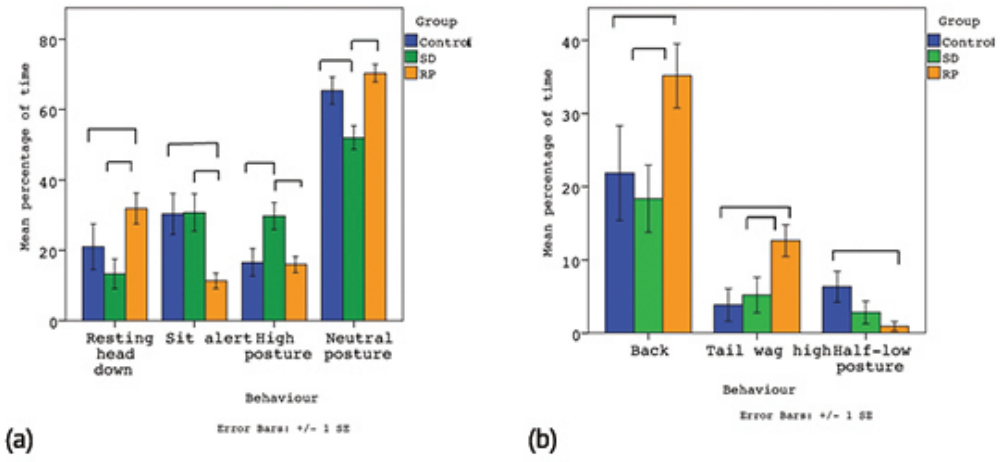


FIGURE 7.4: The significant between-Groups differences in home pen behaviour during the Training Phase

TABLE 7.5: Independent samples t-tests for between-Groups differences in home pen behaviour during the Training Phase

Pair	**Behaviour**	**t(46)**	**p**
Control-SD	High posture	2.402	.020
	Neutral posture	2.628	.012
Control-RP	Sit alert	3.741	<.001
SD-RP	Resting head down	2.705	.009
	Sit alert	4.025	<.001
	High posture	3.253	.002
	Neutral posture	4.316	<.001

TABLE 7.6: Results of Mann-Whitney U tests for between-Groups differences in home pen behaviour during the Training Phase

Pair	Behaviour	U	p
Control-RP	Back	415.5	.048
	High tail wagging	376.5	.006
	Half-low posture	433.0	.003
SD-RP	Back	393.0	.025
	Tail wagging high	405.0	.020

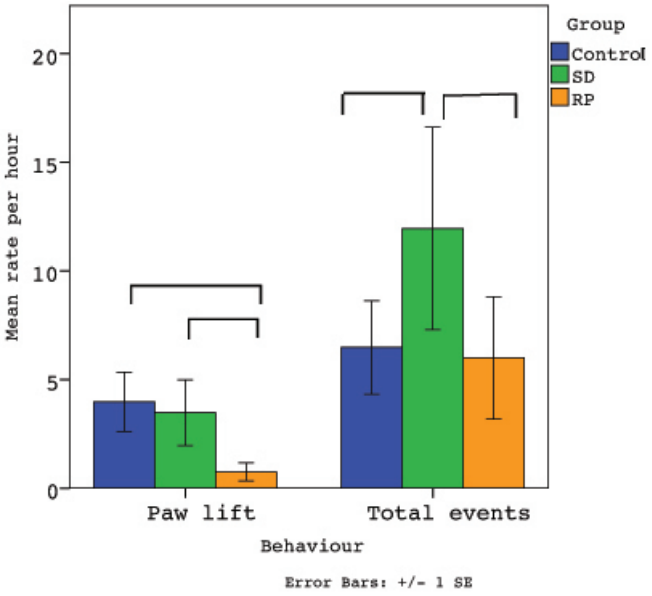


FIGURE 7.5: Behavioural events in the home pen by Group during the Training Phase

The results presented in Tables 7.10 and 7.8 suggest that during the Training Phase, RP dogs were already showing fewer undesirable behaviours and more desirable behaviours. This pattern was also seen in negative behavioural events.

7.3.1.2 Behavioural events

Two behavioural events differed significantly between Groups (Table 7.4). Planned post-hoc Mann-Whitney U tests were conducted to determine where the between-Groups differences lay (Table 7.8).

The results of analyses of behaviour during the Training Phase show several differences in behaviour between Groups. As expected, there were few behavioural differences between Control and SD Groups as they were given identical treatment during this Phase. The only exception is that SD Group spent more time with high posture and

Table 7.7: Behavioural events in the home pen by Group during the Training Phase

Behaviour	χ^2	**p**	**Main findings**
Paw lift	7.052	.029	C,SD>RP
Total events	6.266	.044	SD>C,RP

Table 7.8: Results of Mann-Whitney U Tests between-Groups differences in home pen behaviour during the Training Phase

Pair	**Behaviour**	**U**	**p**
Control-RP	Paw lifts	442.5	.008
	Total events	460.0	.056 (NS)
SD-RP	Paw lifts	489.0	.057 (NS)
	Total events	430.5	.019

Table 7.9: Results of Kruskal-Wallis tests between-Groups differences in home pen behaviour during the SD Phase

Behaviour	χ^2**(2)**	**p**	**Main findings**
Amicable	9.852	.007	RP>C,SD
Play	10.986	.004	RP>SD
High tail wagging	6.503	.039	RP>C,SD
Half-low posture	8.256	.016	SD>RP

less time with neutral posture, however this is not reflected in a difference in behaviour. RP Group spent more time resting head down and at the back of the pen, less time sitting alert, less time with high posture and more time with neutral posture. They also spent more time tail wagging and less time with half-low posture. In RP Group, these changes in behaviour (compared to Control and SD Groups) show increases in desirable behaviours and decreases in undesirable behaviours. As expected, this shows that the training undergone in the Training Phase had a positive effect on welfare.

7.3.2 Between-Groups differences in home pen behaviour during SD Phase

There were between-Groups differences for several behaviours during the SD Phase. There was an effect of Group on sitting alert ($F(2, 68)=3.746$, $p=.029$). There results of Kruskal-Wallis tests are shown in Table 7.9.

Planned post-hoc t-tests and Mann-Whitney U tests were conducted to determine the between-Groups differences. The results are shown in Tables 7.10 and 7.19.

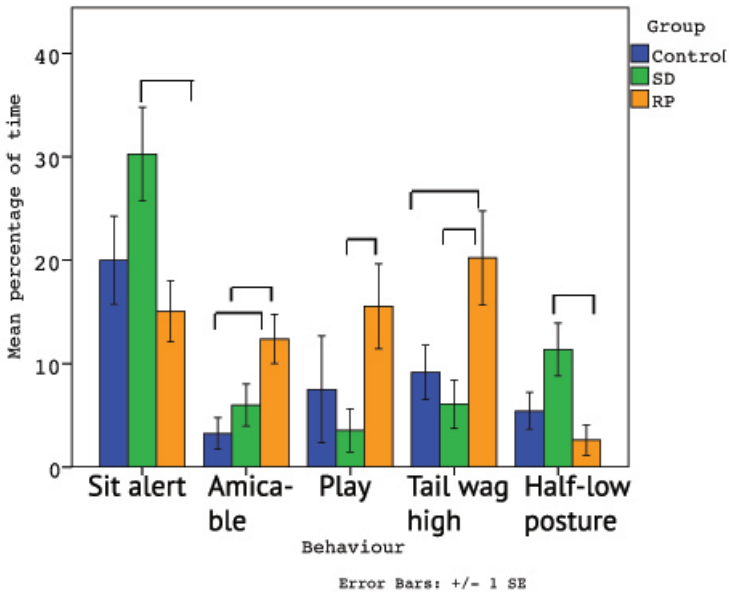


FIGURE 7.6: Between-Groups differences in home pen behaviour during SD Phase

TABLE 7.10: Independent samples t-tests for between-Groups differences in home pen behaviour during the Training Phase

Pair	**Behaviour**	**U**	**p**
Control-SD	Paw lift	189.0	.007
Control-RP	Amicable	154.5	.003
	Play	178.0	.008
	Paw lift	182.0	.009
SD-RP	Amicable	192.0	.004
	Play	177.0	.009
	Tail wagging high	178.5	.006
	Half-low posture	170.5	.006
		t(5)	**p**
	Sitting alert	2.787	.008

There were again few differences between Control and SD Groups, suggesting that SD had had little effect on home pen behaviour for SD Group dogs. RP Group dogs continued to exhibit more desirable behaviours than Control or SD Groups such as play, amicable behaviour, tail wagging, and also exhibited fewer undesirable behaviours such as sitting alert, paw lifts or half-low posture. This suggests that during the SD Phase, the training, which RP Group dogs had undergone in the Training Phase, in combination with the Refined SD procedure, had continued to maintain improved welfare over that of other Groups.

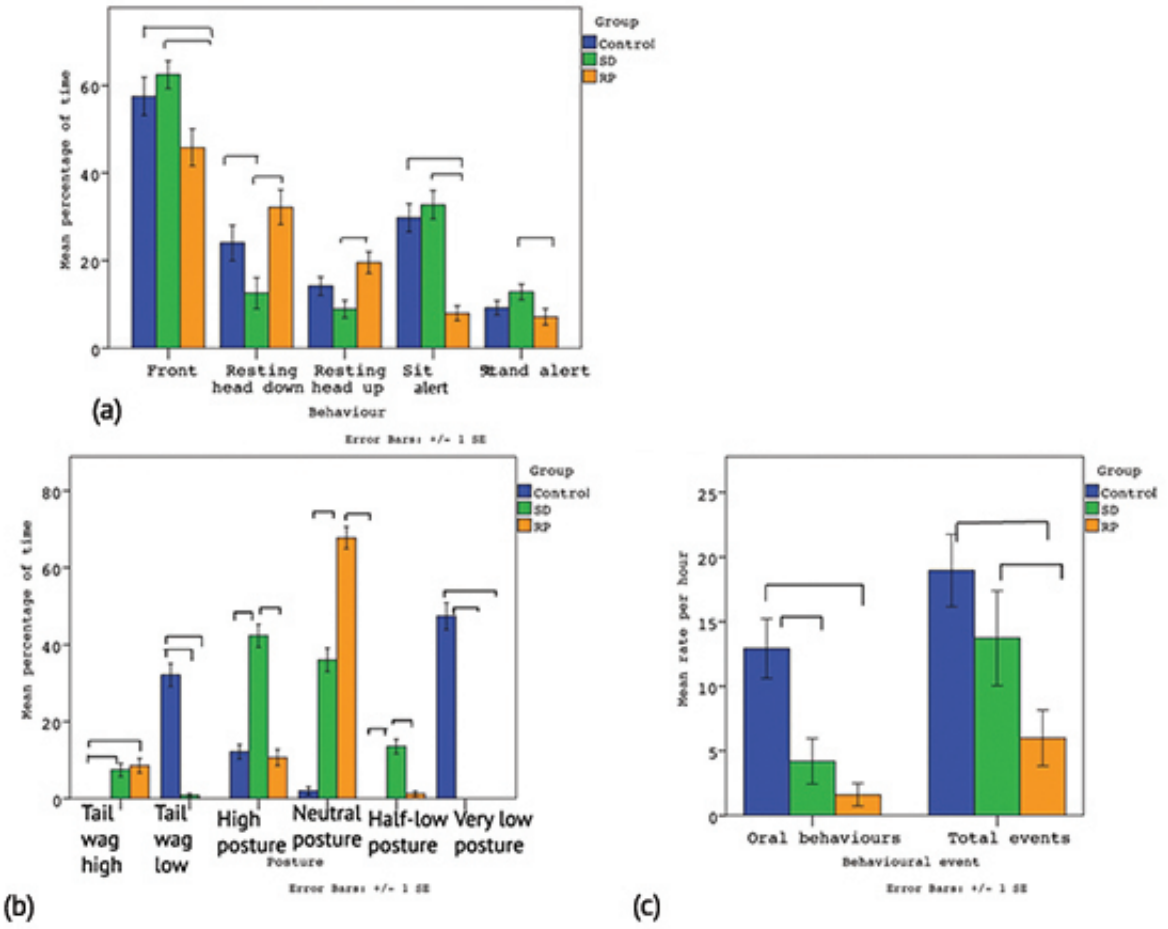


FIGURE 7.7: Between-Groups differences in (a) behavioural states, (b) posture and (c) behavioural events in the home pen during the Dosing Phase

TABLE 7.11: Results of ANOVAs between Groups during the Dosing Phase

Behaviour	**F(2, 165)**	**p**	**Main findings**
Sitting alert	22.616	<.001	C,SD>RP
High posture	35.338	<.001	SD>C, RP
Neutral posture	26.150	<.001	RP,C>SD
Resting head up	5.746	.004	RP>SD

7.3.3 Between-Group differences in home pen behaviour during Dosing Phase

The effects of Group on home pen behaviour were also analysed during the Dosing phase. Significant between-Group differences are shown in Figure 7.7. The results of ANOVAs and Kruskal-Wallis tests are presented in Tables 7.11 and 7.12.

Table 7.12: Results of Kruskal-Wallis tests showing between-Groups differences during the Dosing Phase

Behaviour	χ^2**(2)**	**p**	**Main findings**
Front	7.677	.022	SD>RP
Resting head down	15.276	<.001	C,RP>SD
Standing alert	22.918	.022	SD>RP
Tail wagging high	22.918	<.001	SD,RP>C
Tail wagging low	6.068	.048	SD>C,RP
Half-low posture	61.358	<.001	C,SD>RP
Low posture	6.067	.048	SD>C,RP
Paw lifts	29.508	<.001	C>SD,RP
Total events	18.836	<.001	C,SD>RP

Where significant effects of Group were found, planned post-hoc t-test and Mann-Whitney U tests were conducted to determine where between-Groups differences lay. The results of these are shown in Tables 7.13 and 7.14.

Table 7.13: Results of t-tests showing between-Groups differences in home pen behaviour during the Dosing Phase

Pair	**Behaviour**	**t(118)**	**p**
Control-SD	High posture	8.580	<.001
	Neutral posture	2.433	.016
Control-RP	Sit alert	5.936	<.001
	Neutral posture	4.529	<.001
SD-RP	Sit alert	6.670	<.001
	Resting head up	3.301	.001
	High posture	8.651	<.001
	Neutral posture	7.605	<.001

7.3.4 The effect of time (before/after) on home pen behaviour

Behaviour in the home pen was observed before and after each training or dosing session to determine if there was an immediate effect on behaviour. It was anticipated that there would be increases in undesirable behaviours following sham dosing or dosing events. As such, ANOVAs were conducted with a between-subjects factor of Groups and a within-subjects factor of time (before or after). The results are shown in Table 7.15.

There was no interaction between Group*Time*Phase suggesting that for these behaviours, Phase did not affect any changes in behaviour from before to after training or dosing. There was also a decrease in panting from before to after for all dogs (U=14265.0, p=.006). This may be associated with the decrease in activity.

Table 7.14: Results of Mann-Whitney U tests showing between-Groups differences in home pen behaviour during the Dosing Phase

Pair	**Behaviour**	**U**	**p**
Control-SD	Resting head down	1388.0	.011
	Tail wagging high	1260.0	<.001
	Half-low posture	870.0	<.001
	Paw lifts	2286.0	.002
	Total events	1363.0	.013
Control-RP	Front	1414.5	.040
	Tail wagging high	1230.0	<.001
	Half-low posture	1054.	.043
	Paw lifts	1002.5	<.001
	Total events	1080.0	<.001
SD-RP	Front	1294.5	.007
	Resting head down	1139.5	<.001
	Standing alert	1353.0	.007
	Half-low posture	958.0	<.001
	Paw lifts	1540.0	.028

Table 7.15: Results of ANOVAs showing effects of Group and time (before or after) on home pen behaviour

Effect or interaction	**Behaviour**	**F**	**df**	**p**	**Main findings**
Time	Resting head down	4.568	1,116	.03	After>Before
	Play	16.149	1, 116	<.001	Before>After
Group∗Time	Resting head up	4.144	2, 116	.025	Control After>Before
	Back	3.716	2, 116	.025	Control After >Before
	High posture	4.458	2, 116	.012	RP Before >After
	Neutral posture	3.784	2, 116	.025	RP After>Before

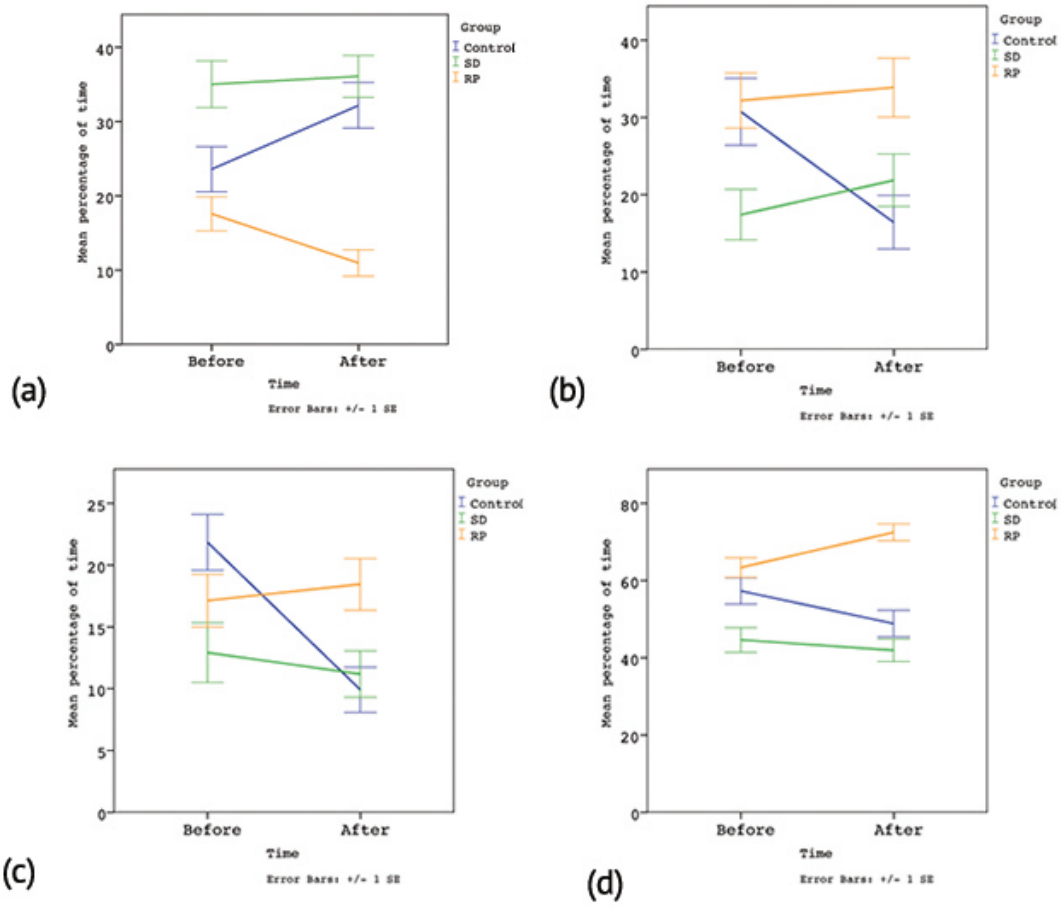


FIGURE 7.8: Mean percentage of time (before/after) spent at (a) back, (b) resting head up, (c) high posture and (d) neutral posture, by Group

Control Group spend less time resting and at the back of the pen after a training, SD or dosing session. This suggests that either the interventions given to SD and RP Groups caused them to be less active following SD or dosing sessions, or that Control Group dogs were more active due to activity within the unit or lack of predictability regarding being removed from the pen for dosing. RP Group decreased the percentage of time spent with high posture after, and also increased time with neutral posture. The increase in high posture for Control Group was marginally non-significant (p=.05).

TABLE 7.16: Results of t-tests showing significant changes in home pen behaviour over time (before or after) by Group

Group	**Behaviour**	**t**	**df**	**p**	**Main findings**
Control	Back	2.579	106	.011	Before>After
	Resting head up	4.125	106	<.001	Before>After
RP	High posture	2.278	129	.024	After>Before
	Neutral posture	2.687	129	.008	Before>After

SD Group showed no significant changes in behaviour over time. When non-parametric data were split by Group, there was no significant change in any behaviour over time for Control or SD Groups (.921<p<1.000), however RP Group showed a decrease in rapid locomotion (U=1947.0, p=.012) and panting (U=1947.0, p=.012) from before to after training/dosing sessions.

7.3.5 The effects of Phase on home pen behaviour

It was anticipated that behaviour would change across successive Phases. To investigate the effects of Phase on home pen behaviour, ANOVAs and Kruskal-Wallis tests were conducted. The results are shown in Tables 7.17 and 7.18.

TABLE 7.17: Results of ANOVAs showing effects and interactions of Phase and Group on home pen behaviour

Group	Behaviour	F	df	p	Main findings
Phase	Interact with environment	6.360	2	.002	SD,Dosing>Training
	High posture	4.154	2, 338	.012	Dosing>Training
	Neutral posture	74.161	2, 338	<.001	Training>Dosing
	Play	5.085	2, 338	.007	Training,SD>Dosing
Group*Phase	High posture	9.178	4, 338	<.001	SD Group: SD >Dosing Phases
	Neutral posture	2.940	4, 338	.021	SD Group: SD>Dosing Phases

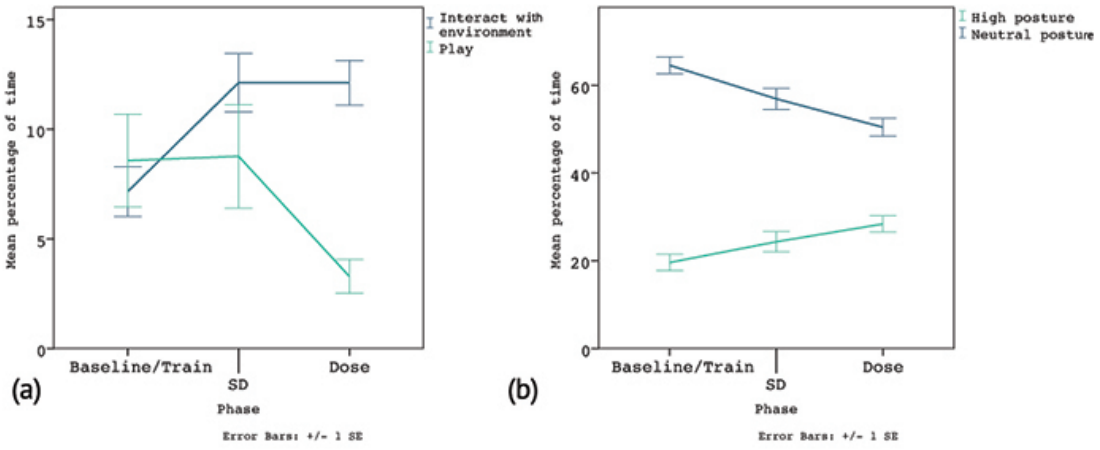


FIGURE 7.9: The effects of Phase on (a) interaction with environment and play and (b) high and neutral posture in the home pen

Interacting with the environment increased from the SD Phase to the Dosing Phase, while play behaviour decreased. Neutral posture decreased in the SD Phase and again in the Dosing Phase. These behaviours have been affected by the doses administered in the Sham Dosing and Dosing Phases.

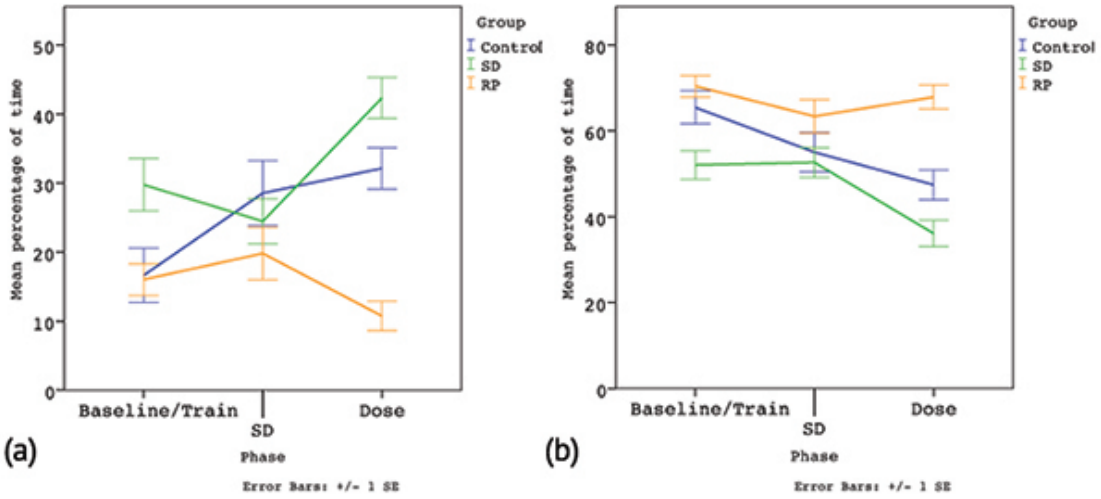


FIGURE 7.10: The interaction between Group and Phase on (a) high posture and (b) neutral posture in the home pen

Control and RP Groups showed similar pattern in high posture over time, while SD Group showed an increase between the SD and Dosing Phases. Neutral posture did not change over time for RP Group, while both Control and SD Groups showed a decrease during the Dosing Phase.

TABLE 7.18: Results of Kruskal-Wallis tests showing effects of Phase on home pen behaviour

Behaviour	χ^2**(2)**	**p**	**Main findings**
Calm locomotion	6.128	.047	Decrease during Dosing
Tail wagging	10.98	.004	Decrease: SD < Dosing
Half-low posture	17.160	<.001	Increase: Training<SD<Dosing
Paw lifts	6.291	.043	Increase: Training < Dosing
Total events	6.850	.033	Increase: Training <SD

There was a marginally non-significant effect of Phase on Resting head up (.053), due to a decrease from the Training to SD Phases, and a marginally non-significant interaction (p=.069) on time at barrier due to SD Group, but not the other groups, increasing time at barrier in the the SD Phase.

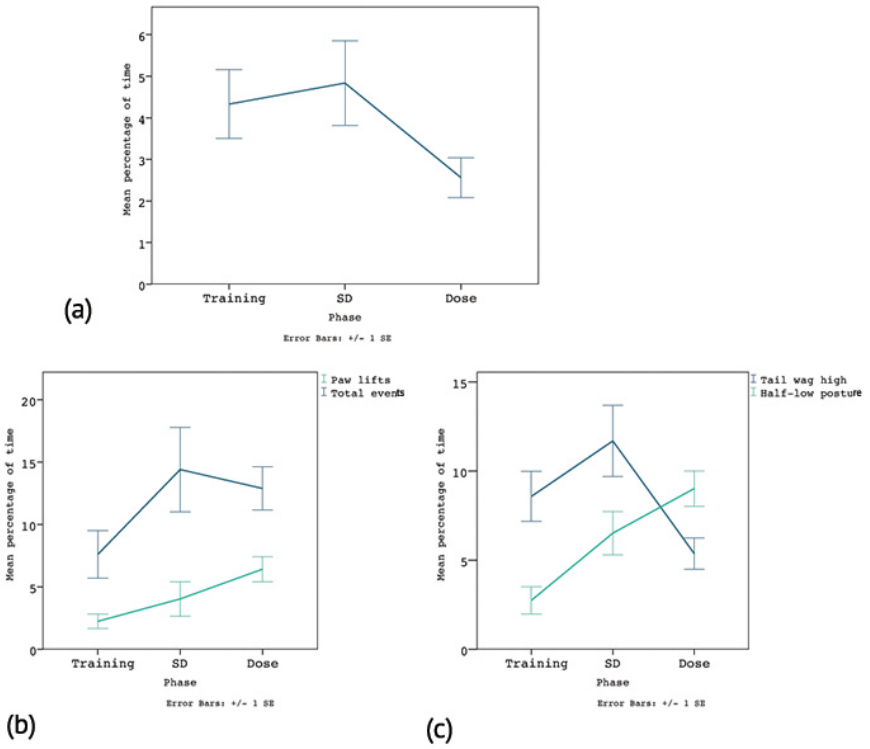


FIGURE 7.11: Effect of Phase on (a) calm locomotion, (b) behavioural events in the home pen and (c) postures

Calm locomotion decreased in the Dosing Phase as did high tail wagging while low tail wagging, which was rarely seen before, increased. Similarly, half-low posture increased in the SD Phase, while very low posture was only seen in the Dosing Phase. Paw lifts increased in the Dosing Phase while all events increased from the SD Phase.

As expected, positive welfare indicators such as calm locomotion and neutral posture decreased across Phases while negative welfare indicators such as high posture and behavioural events increased. The increase in interaction with the environment, which is a positive welfare indicator, may be a response to release from single housing, previously seen in dogs subject to a one-hour single-housing (see Chapter 6).

7.3.6 Behaviour during Sham Dosing

To determine immediate behavioural responses to sham dosing, behaviour was recorded for the duration of the event. Behaviour during SD was scored on a continuous basis, with behavioural states presented as a percentage of time and behavioural events presented as a rate per hour. As expected when only a small number of observations were conducted, data were non-normally distributed. Mann-Whitney U tests were conducted with one between-subjects factor of Group and one within-subjects factor of

Day. Only dogs in SD and RP Groups were sham dosed, with dogs in RP Group receiving a Refined technique. The same protocol for each group of dogs was followed on both days on sham dosing. Mann-Whitney U tests revealed between-Groups differences, as well as one change between days. These are shown in Table 7.19.

TABLE 7.19: Results of Mann-Whitney U tests showing effects of Group and Day on behaviour during sham dosing

Effect	Behaviour	U	p	Main findings
Group	Sit relaxed	7.465	.006	RP>SD
	Collapsed	5.948	.015	SD>RP
	Interact with handler	18.102	<.001	SD>RP
	Interact with environment	4.509	.034	RP>SD
	Freeze	8.492	.004	SD>RP
	Low posture	11.538	.001	SD>RP
	High tail wagging	4.045	.044	RP>SD
Day	Sit relaxed	36.0	.039	RP>SD

There were few differences between the first (SD1) and second (SD2) sham doses. However, it was found that SD Group spent more time exhibiting freeze during SD2 (U=5.5, p=.041) while RP Group spent less time sitting relaxed (U=0.0, p=.002). Sitting relaxed significantly decreased between doses. Between-Groups differences are shown in Figure 7.12.

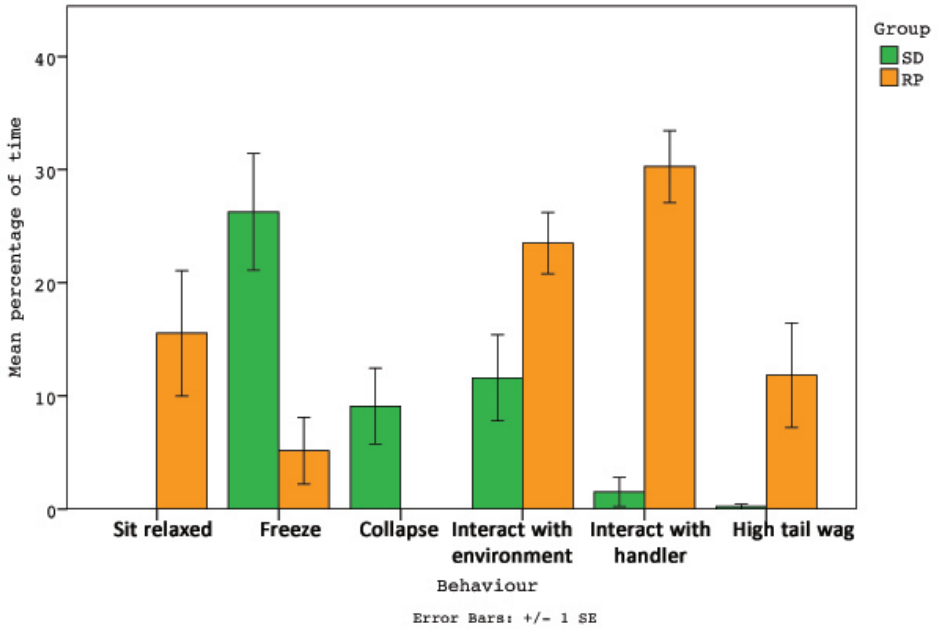


FIGURE 7.12: Significant between-Groups differences in behaviours during Sham Dosing

As expected, RP Group spent more time interacting with the hander, sitting relaxed and less time freezing during sham dosing, suggesting that dosing and training had sufficiently desensitised them to the presence of the handler, technician and equipment before sham dosing began. It is not surprising that sitting relaxed decreased for RP Group and freezing increased for SD Group across doses as the first SD likely had primed a response to the aversive event. It is likely that RP Group exhibited more interacting with environment and high tail wagging due to the positive associations created with the environment and handler during training, over-riding this brief aversive event.

7.3.7 Behaviour during dosing

As with sham dosing, behaviour during dosing was scored on a continuous basis for all Groups, with behavioural states presented as a percentage of time and behavioural events presented as a rate per hour. ANOVAs were conducted with one between-subjects factor of Group and one within-subjects factor of Day. Non-normally distributed data were analysed using a Kruskal-Wallis test. Only four of the five days of dosing had data for every dog, therefore only four days were analysed. Data for some dogs was lost on the fourth day of dosing (Day 18) due to video failure, the partial data were not included.

Table 7.20: Results of ANOVAs showing effects of Group and day on behaviour during dosing

Effect	Behaviour	F	df	p	Main findings
Group	Interact with handler	3.159	2,60	.036	RP>C, SD
	Sit	8.267	2, 60	.001	C,SD>RP
	Struggle	4.523	2, 60	.015	C,SD>RP
	Freeze	27.407	2, 60	<.001	C,SD>RP
	Paw lifts	2.572	2, 60	.034	C,SD>RP
Day	Interact with environment	2.825	3, 60	.046	

There were effects of Group on several behaviours. There were no interactions between Group*Day. Planned post-hoc t-tests were conducted to determine where between-Groups differences lay. The results are shown in Table 7.21. There were no significant differences in behaviour between Control and SD Groups.

As can be seen in Figure 7.13, RP Group spent less time than Control Group or SD Group sitting, more time interacting with the handler, less time struggling, less time 'freezing'; and as can be seen in Figure 7.14, displayed fewer paw lifts than Control Group.

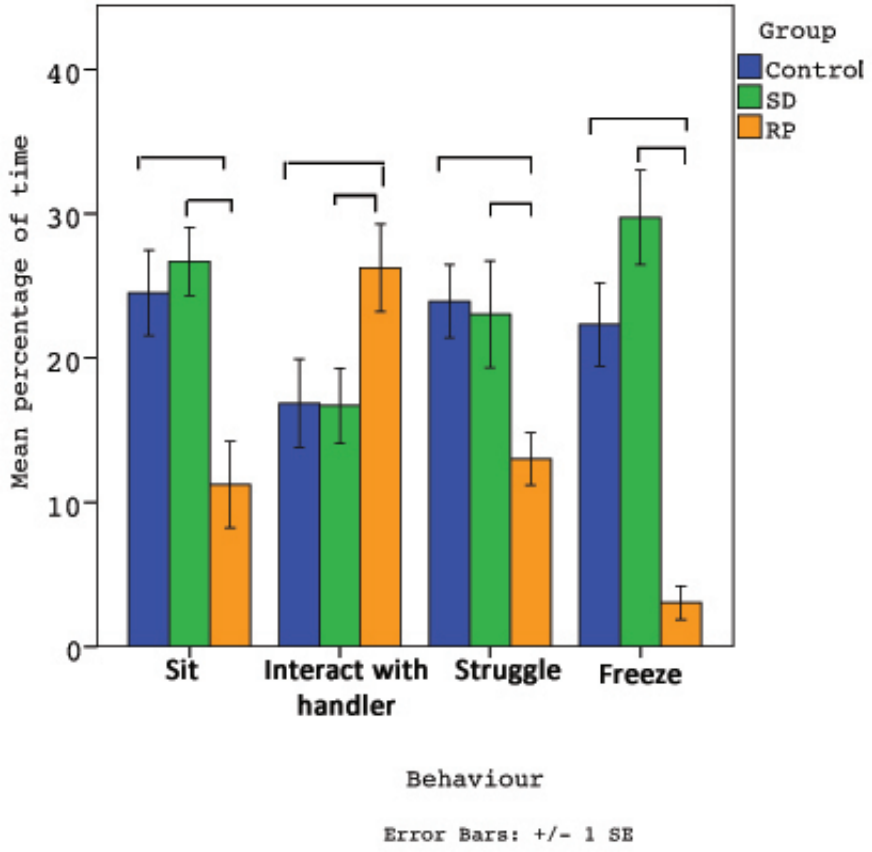


FIGURE 7.13: Mean percentage of time spent in behaviours during Dosing, by Group

TABLE 7.21: Independent samples t-tests for between-Groups differences in behaviour during dosing

Pair	**Behaviour**	**t(46)**	**p**
Control-RP	Interact with handler	2.184	.034
	Sit	3.154	.003
	Struggle	3.509	.001
	Freeze	6.181	<.001
	Paw lifts	3.080	.003
SD-RP	Interact with handler	2.384	.021
	Sit	4.056	<.001
	Struggle	2.423	.019
	Freeze	7.701	<.001

There were effects of Group on several behaviours. There were no effects of Day. Planned post-hoc t tests were conducted to determine where between-Groups differences lay. The results are shown in Table 7.21.

As can be seen in Figure 7.13, RP Group spent more time sitting relaxed than either Control Group or SD Group, less time with low posture, and less time crouching or

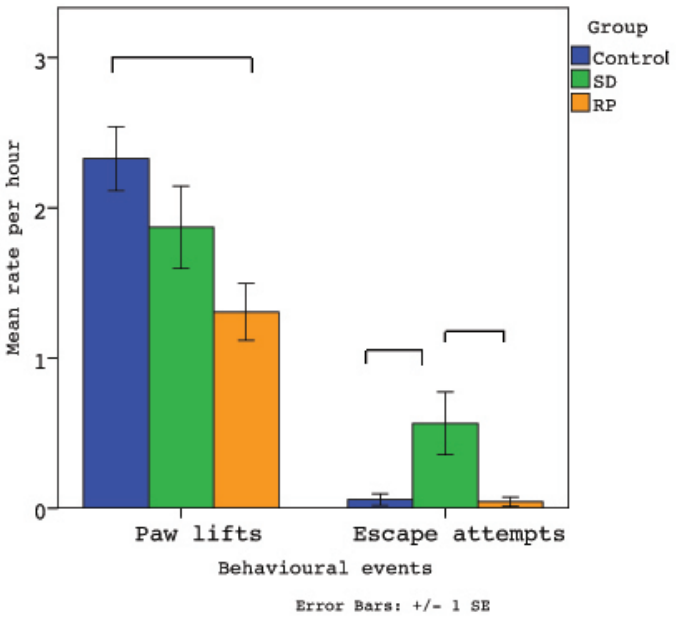


FIGURE 7.14: Mean rate per hour of behavioural events during Dosing, by Group

TABLE 7.22: Results of Kruskal-Wallis tests showing effects of Group on behaviour during Dosing

Behaviour	χ^2**(2)**	**p**	**Main findings**
Sit relaxed	42.751	<.001	RP>C,SD
Stand	9.093	.011	C>RP
High posture	9.755	.008	C,SD>RP
Low posture	29.910	<.001	C,SD>RP
Neutral posture	32.986	<.001	RP>C,SD
Crouch	16.461	<.001	SD>C>RP
Tremble	12.731	.002	C,SD>RP
Escape attempts	7.847	.020	SD>C,RP

trembling. RP Group also spent less time with high posture than SD Group and exhibited fewer escape attempts. There were few differences between Control and SD Groups, however Control Group spent less time crouching than SD Group and made fewer escape attempts.

The differences between RP Group and Control and SD Groups were fairly consistent, suggesting a reliable improvement in welfare regardless of whether compared to sham dosed or control dogs. As with sham dosing, RP Group displayed more behaviours indicative of successful desensitisation such as sitting relaxed, rather than sitting while resisting, freezing or crouching. The rate of oral behaviours was very high in all Groups, which likely reflects anticipatory behaviour as well as a response to dosing. RP Group also spent more time interacting with the handler, which is unsurprising

given that the six sessions of training or sham dosing involved positive interactions with the handler in the same setting as well as the handler being a source of positive reinforcement.

7.3.8 Welfare monitoring tool (WMT) Scores

The WMT was employed to determine if changes in welfare could be detected by the technician in a manner practical to use in the busy laboratory environment. WMT sheets were scored by the technician for each dog on Days 3 and 4 in the Training Phase, Days 9-11 in the SD Phase and Days 15-19 in the Dosing Phase. Behaviour was scored hourly between 8am - 3pm, with the exception of 1pm which was during feeding time. Additional behaviours were scored where dosing or sham dosing took place in any given hour. A low score represented good welfare, with increasing scores representing negative welfare. Sham dosing took place on Days 10 and 11, dosing took place on Days 15-19. Scores are presented as HP (home pen scores only), Dose (score during dosing only) and combined scores. The effects of Group and time on these scores were analysed using an ANOVA with one between-subjects factor of Group and one within-subjects factor of Day.

7.3.8.1 Home pen only

There was a significant effect of day on score, $F(9, 169)=2.385$, $p=.014$, due to a decrease from Days 4-11 ($t(17)=2.858$, $p=.007$), an increase from Day 11-15 ($t(17)=2.159$, $p=.038$) and decrease from day 15-17($t(17)=2.429$, $p=.021$).

There was a marginally non-significant effect of Group on HP score ($p=.055$). This can be seen in Fig 7.16 with RP Group showing a trend towards lower HP score than SD Group.

There was an effect of Phase on HP score ($F(2, 170)=3.394$, $p=.036$). This was due to scores being lower in the Dosing Phase than in the Training Phase (see Figure 7.17).

There was also a significant interaction between Group and Phase ($F(4, 170)=2.502$, $p=.044$). In the Training Phase, SD Group had a high score than Control or SD Groups ($F(2, 170)=3.820$, $p=.029$) and in the Dosing Phase, RP Group had a lower score than Control or SD Groups ($F(2, 170)=3.374$, $p=.039$).

As Figure 7.18 shows, while the WMT score for Control Group changed little across the study, the score for RP Group decreased. The error bars also suggest that there was less variation in the behaviour of the RP Group dogs than others.

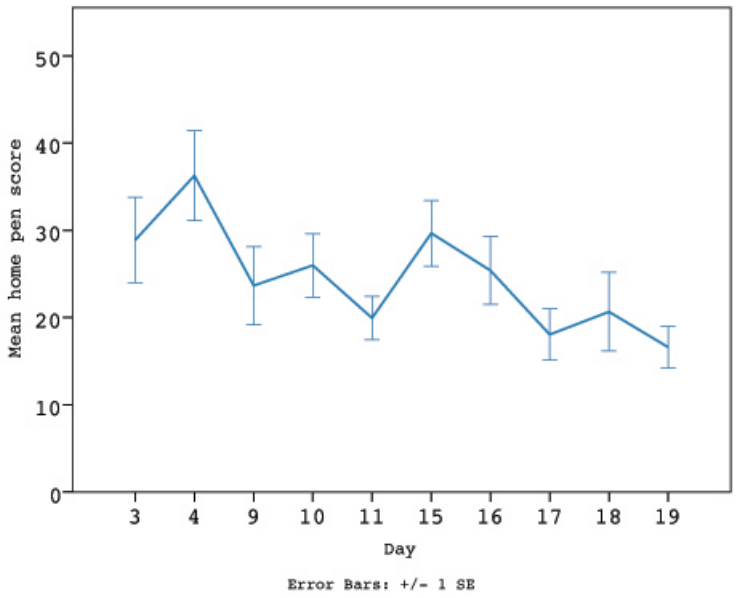


FIGURE 7.15: Mean WMT Home Pen scores across all days for all dogs (of maximum 40)

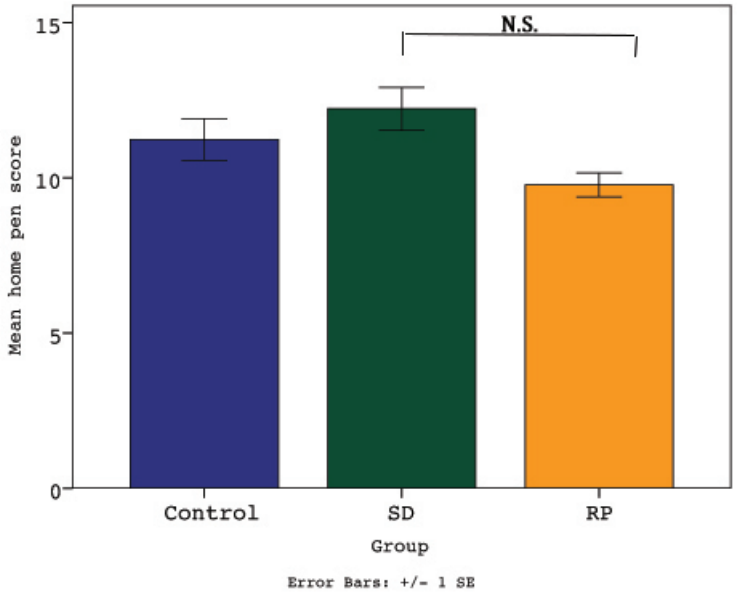


FIGURE 7.16: Mean WMT Home Pen by Group (of maximum 40)

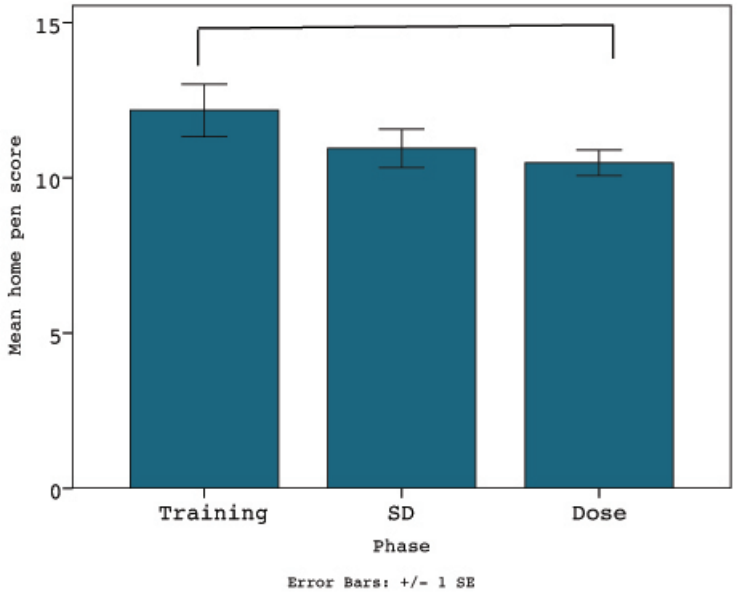


FIGURE 7.17: Mean WMT Home Pen by Phase (of maximum 40)

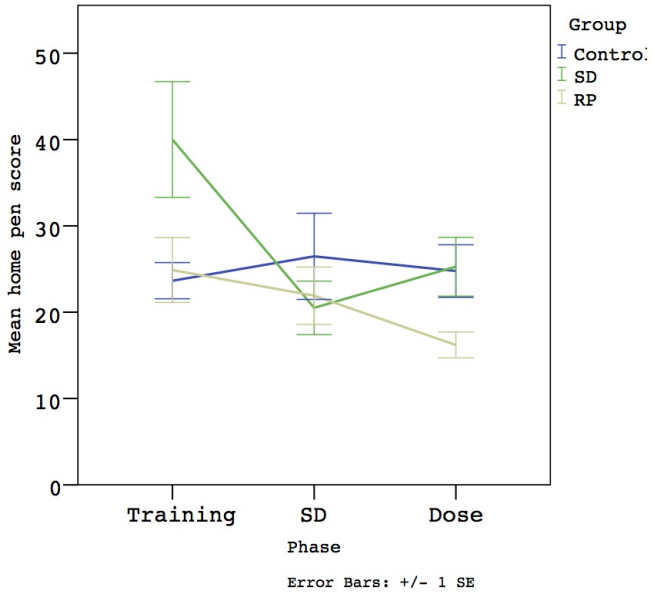


FIGURE 7.18: Mean WMT Home Pen score across Phases and between Groups (of maximum 40)

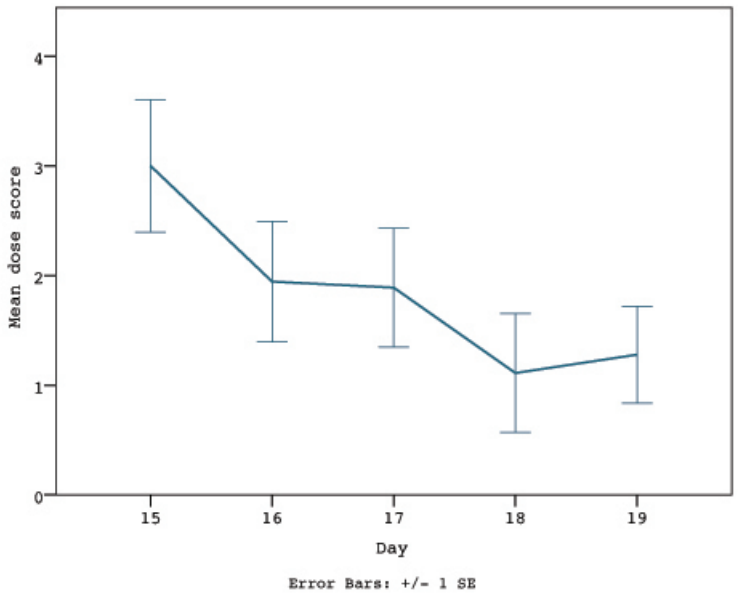


FIGURE 7.19: Mean WMT dosing score out of a possible 11 across all five dosing days

7.3.8.2 Dosing scores only

There was a significant effect of day on score, F(4, 85)=3.183, p=.019 due to a significant decrease between Day 15-18 and Day 15-19. There was no interaction between Day*Group (p=.107).

TABLE 7.23: Paired-samples t-tests for WMT dosing score between days, showing significant differences

Pair	**t(17)**	**p**
Day 15-18	3.019	.008
Day 15-19	3.144	.006

7.3.8.3 Total score (Home pen + dosing)

There was a significant effect of day on score (F(9, 163)=2.162, p=.027). As with HP score, there was a significant decrease from Days 4-11 (t(17)=2.238, p=.033). There was also a significant increase between Day 11-15 (t(17)=2.166 p=.039) and decrease between Day 15-17 (t(17)=2.808, p=.012).

There was an effect of Group on total score (F(2, 163)=2.502, p=.044), with a similar trend seen as in the HP scores. RP Group had a lower score than SD Group (t(118)=1.875, p=. 005). The difference between Control and RP Groups was non-significant (p=.063). There was no interaction between Group*Phase (p=.094).

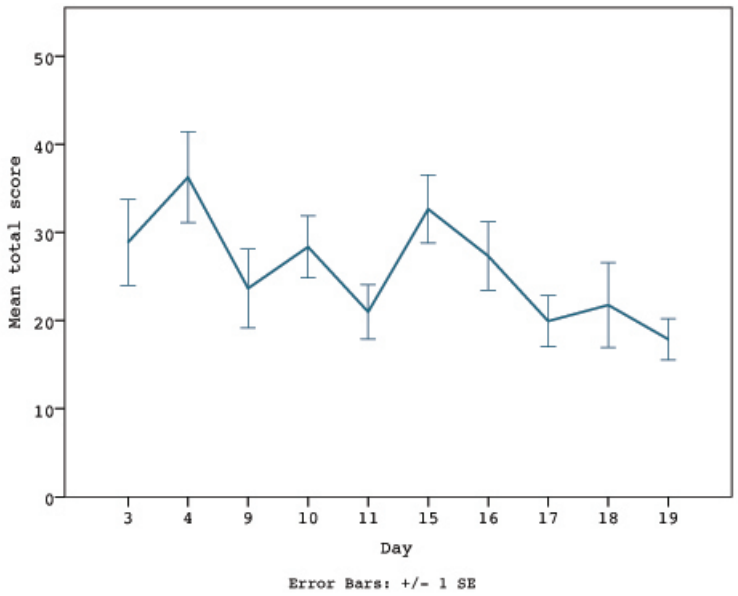


FIGURE 7.20: Mean WMT total score for all dogs across all days (of maximum 51)

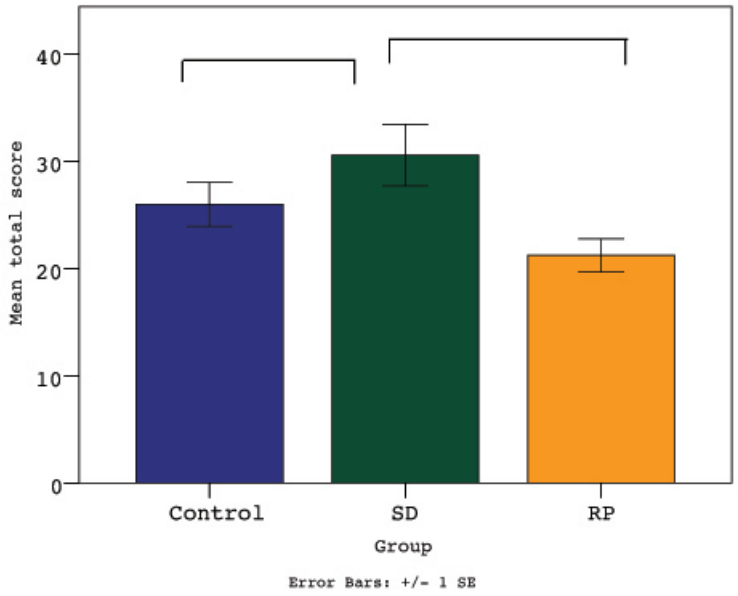


FIGURE 7.21: Mean WMT total score by Group for all days (of maximum 51)

Scores follow the expected pattern in that there was a significant increase in home pen scores on Day 15, the first day of dosing, which had not decreased to pre-dosing levels until Day 17. Dosing scores also decreased over the Dosing Phase, reflecting an increasing habituation to dosing. The between-Groups differences in home pen behaviour are reflected in WMT scores, with significant differences in score between SD and RP Groups, with marginally non-significant differences with Control Group.

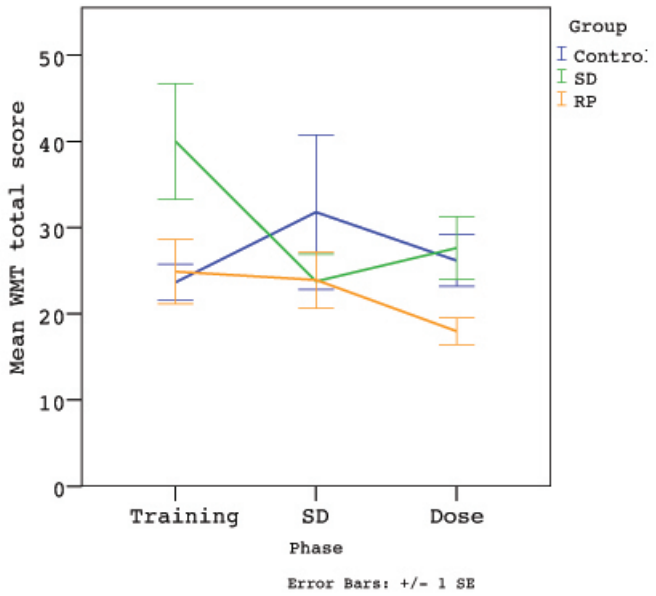


FIGURE 7.22: Mean WMT total score by Group across Phases (of maximum 51)

7.3.9 Mechanical pressure threshold (MPT) testing

MPT readings (N) were taken on five days in total: Days 10 and 11 in the SD Phase, and Days 15, 17 and 19 in the Dosing Phase. Technical issues meant it was not possible to take readings during the Training Phase. Readings taken in SD and the Dosing Phases allow a comparison of MPT change as a result of dosing. Three readings were taken on each day, with the mean calculated from these readings. There was a significant effect of Day on MPT, $F(4, 160)=9.622$, $p<.001$. MPT dropped from Day 10-11, and again from 11-15. There was no significant change from Day 17-19 ($p=.641$), and there was a non-significant difference between scores on day 15 and Day 19 ($p=.086$).

However, there was a significant interaction between Group*Time, $F(8, 160)=4.589$, $p<.001$. This is due to RP Group showing no change over time ($p=.149$), while Control Group ($F(4, 20)=10.17$, $p<.001$) and SD Group ($F(4, 20)=137.29$, $p<.001$) did. This contrast between RP Group and Control and SD Groups suggests that RP Group were less susceptible to the change in MPT caused by SD or dosing. The Welfare Assessment Framework (Chapters 4-6) suggests that this reflects a lack of change in affective state.

The decrease in MPT between Days 10-11 most likely reflects a change caused by SD, as does the further decrease from Day 10-15, with dosing. It was expected that events which caused a change in affective state would cause a change MPT and as the first SD (for SD and RP Groups) and the first dose (for all Groups) were two of the most

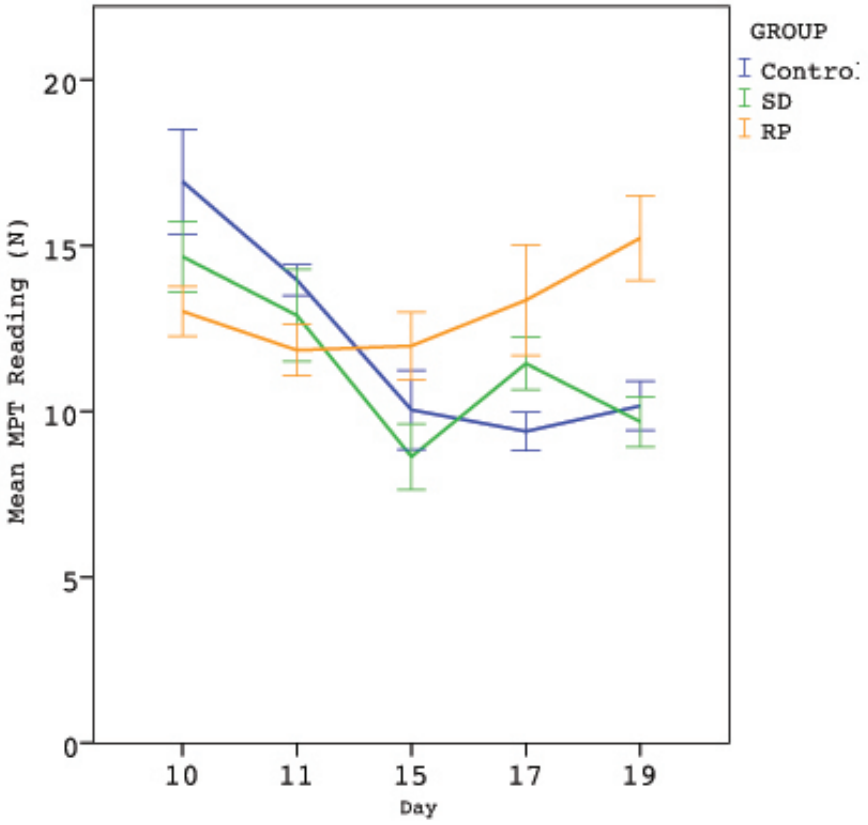


FIGURE 7.23: MPT readings over five days of dosing by Group

TABLE 7.24: Paired-samples t-tests for MPT readings

Pair	t(17)	p
Day 10-11	2.202	.042
Day 11-15	2.608	.018

aversive events during the study (being the least predictable), the change in MPT reflects this. RP Group did not show the same changes in MPT over time.

7.3.10 Visual analogue scales

VAS scales (0-10 cm line) were used in the Dosing Phase to score the technician's perceived overall ease of dosing for each dog on each of the five dosing days. The technician scored each dog once immediately following each dosing session. There was a significant effect of Day on score, $F(4, 60)=5.529$, $p<.001$. This was the result of an increase between Days 15-18, Days 16-19 and an overall increase from Day 15 Day 19 (Table 7.25). There was no interaction between Group*Day ($p=.902$). This suggests

that dogs became easier to dose with successive doses, regardless of pre-study treatment.

TABLE 7.25: Paired-samples t-tests for showing significant effects of Day on VAS

Pair	**t(46)**	**p**
Day 15-18	4.320	<.001
Day 15-19	5.476	<.001
Day 17-19	2.272	.036

VAS scores increased across the Dosing Phase, which as expected shows that dogs became easier to dose with each successive dose. This agrees with the decrease in dose score from the welfare monitoring tool over dosing days. There is no clear pattern between-Groups, perhaps as the result of the large amount of within-Groups variation evident. It is worth noting that the factors noted by the technician as contributing most to ease of dosing included the size of the dog, size and shape of the muzzle as well as behaviour. It is reasonable to assume that the between-Groups differences seen in Section 1.3.8 are not reflected in these scores as they take into account factors other than behaviour. No guidance relating to specific parameters was given to the animal technician on scoring dogs using the VAS in order to prevent bias, with the exception being to score "ease of dosing". As ease of dosing is influence by factors other than behaviour, it is possible that a different measure needs to be developed to monitor the effects of training on technician satisfaction with dogs during dosing, or that in a short-term study such as this, technician satisfaction is not one of the more important factors to consider when welfare and time to dose are positively influenced by Refined protocols.

7.3.11 Food consumption

Food consumption (FC) was measured daily from Day 1 to Day 19, as a weight in grams eaten from a 300g ration. Data from Day 13 (a Saturday) were discarded, as food consumption for all dogs dropped significantly for reasons unlikely to be related to experimental protocols and likely to be due to an unfamiliar member of staff conducting cleaning and feeding duties on that day. Data were analysed using a repeated measures ANOVA to investigate the effects of Group and Day on food consumption, and the interactions between these factors. The results of these analyses are shown in Table 7.26 and show that food consumption did vary between days, and that there was a significant interaction between Group and Day.

As Fig 7.26 shows, there are several significant differences in FC across days and t-tests were conducted to determine where significant differences lay. The highest values shown

TABLE 7.26: Results of repeated measures ANOVA for food consumption

Effect or interaction	F	df	p	Difference
Day	3968	17, 255	<.001	multiple
Day∗Group	1.756	34, 255	.008	multiple

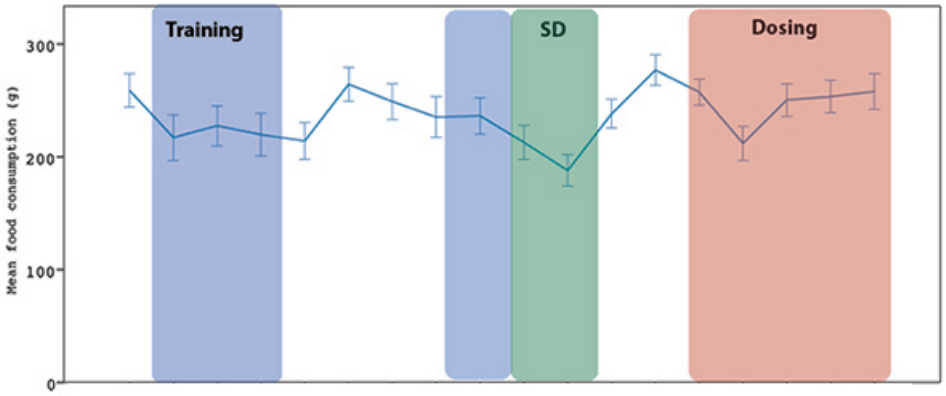


FIGURE 7.26: Food consumption (of 300g) for all dogs across all days

in the figure are Saturdays, with the second highest being Sundays.. As FC data were not available before Day 1, it is not possible to say if there was a drop in FC on Day 2, or if it was higher than usual on Day 1, but as there was a significant drop (p=.020) followed by a significant increase from Days 5-6 (p=.006), it would seem that FC dropped from the first day of training. There was no significant drop in FC on the final day of training, or first day of SD, but FC on the next day was significantly lower (p=.008). There was no significant decrease in FC during the Dosing Phase, with the exception of the second day6 (p=.003).

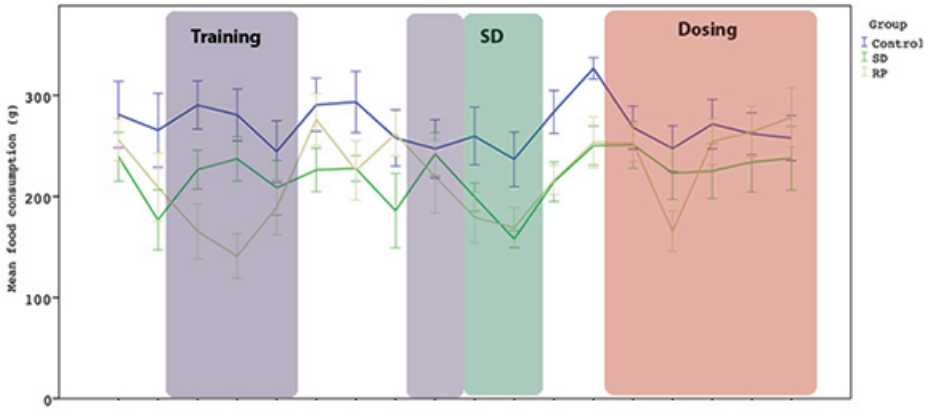


FIGURE 7.27: Food consumption (of 300g) by Group across all days

The mean FC for each of the three Phases was calculated for each Group to determine if Phase affected Groups' FC differently. These means are shown in Figure 7.26. Mean

food consumption between Groups across all Phases, with the results of between-Groups ANOVAs shown in Table 7.27. Food consumption varied between-Groups in the Training Phase and the SD Phase as the result of Control Group having higher FC than SD or RP Groups, which were not significantly different. There was no significant difference in FC between-Groups during the Dosing Phase (p=.657) or for all dogs across Phase (p=.130).

TABLE 7.27: Results of ANOVAs for FC between Groups across Phases

Phase	**F(2, 15)**	**p**	**Main findings**
Training	4.194	.036	C>SD,RP
SD	6.469	.009	C>SD>RP

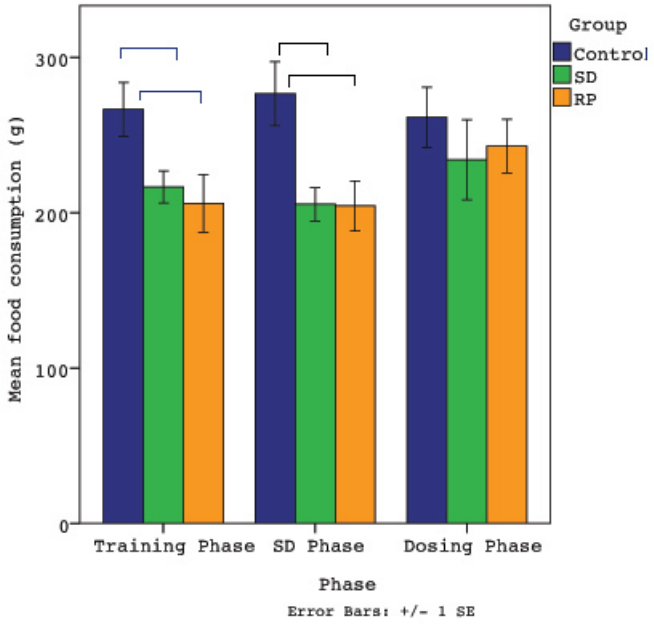


FIGURE 7.28: Mean food consumption between Groups across all Phases

This pattern of food consumption seems to be consistent across Phases. Without pre-study food consumption, it is not possible to say if Control Group always has higher food consumption than the other Groups, but the data presented in Fig 7.28 suggests this is consistent.

7.3.12 Body weight

Body weight was measured once weekly for each dog from Week -1, the week before the Training Phase began. A repeated-measures ANOVA was conducted with one within-subjects factor of Week (4 levels) and one between-subjects factors of Group (3 levels). There was no significant effect of Week (p=.532), nor a significant interaction between

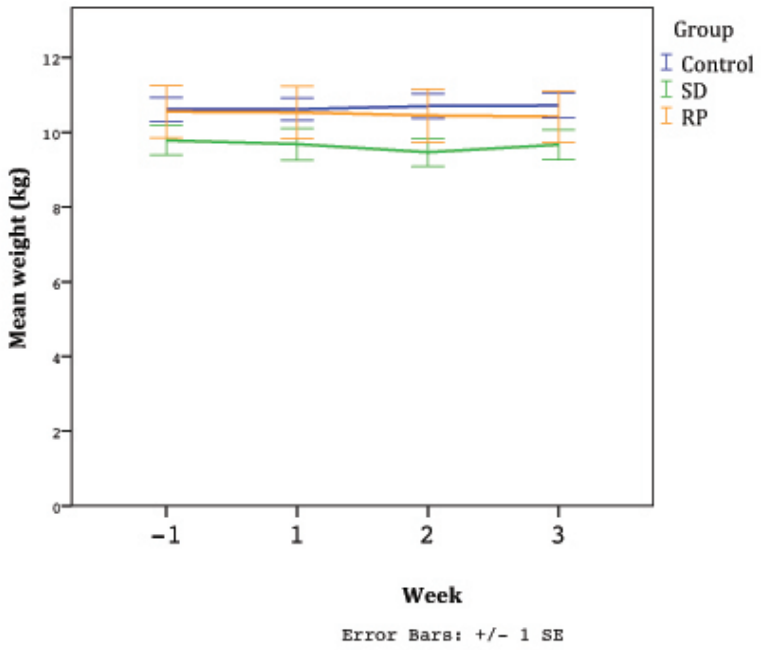


FIGURE 7.29: Mean body weight by Group, across all Weeks

Group*Week (p=.397). As Figure 7.29 shows, body weight was very stable across the study regardless of Group.

7.3.13 Time to dose

One of the aims of improving the behaviour and cooperation of the dogs during dosing was to improve the ease and speed of protocol for the technician. As such, the time taken to dose each dog was measured during each dose. This included time to restrain the dog to attach the heart rate strap, restrain in the dosing position, dose, remove the heart rate strap and lift the dog from the table. Data from Dose 4 were again discounted due to missing values.

The range of times taken for dosing was 46.02 - 1:43.04. There was a significant effect of Dose ($F(3, 60)=10.668$, $p<.001$) on time taken, shown in Fig 7.28, with Dose 1 taking longer than Doses 2 (p=.003), 3 (p=.004) and 5 (p<.001).

In addition, time to dose was compared between Groups. While the same pattern of decreasing time across doses was maintained, there was an effect of Group ($F(2, 60)=2.317$, p=.025), with SD taking significantly longer to dose than Control ($t(46)=2.249$, p.029) and RP ($t(46)=2.054$, p=.046). There was no difference between Control and RP (p=.9).

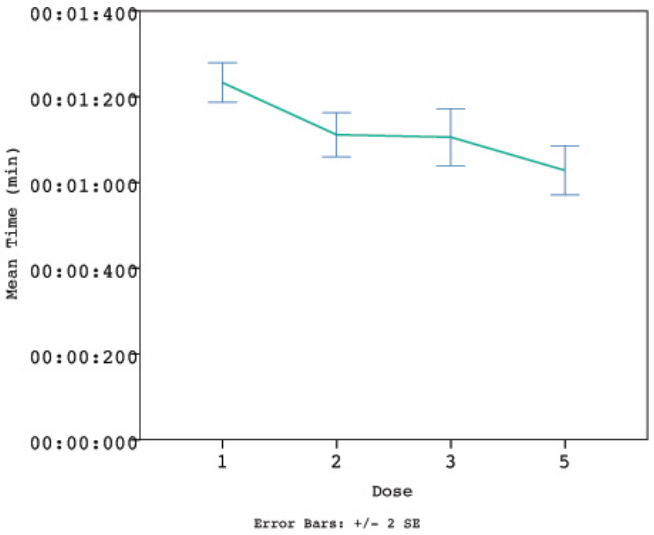


FIGURE 7.30: Mean time to dose by day, collapsed across all Groups

While RP Group did not prove to be the quickest Group to dose, dosing times were improved over the SD Group. There was no interaction between Group*Dose. This should be taken into consideration when weighing up the benefits of pre-study training protocols.

7.4 General Discussion

The first aim of this study was to compare a number of variables between two Groups of dogs subject to oral gavage, one of which was given sham dosing treatment pre-study (SD Group) and one which was given no treatment pre-study (Control Group) to determine if there was a benefit of the sham dosing procedure to welfare. A further aim was to compare these two Groups with a third which received a Refined training protocol pre-study and Refined handling during dosing (RP Group) to determine if the treatment and time investment resulted in significant benefits to welfare.

7.4.1 Behaviour in the home pen and during dosing

During the Training Phase, Control and SD Groups were given no interventions and as expected there were few differences in behaviour between them. There were a number of differences with RP Group however. There were differences key behaviours such as resting head down, sitting alert, high posture, neutral posture and behavioural events, with the overall pattern showing that RP Group dogs were less vigilant and more restful in the home pen. As during the Training Phase, RP Group were given four training

sessions, incorporating PRT, Refined handling and predictability but no aversive events it is not surprising that the dogs' welfare was higher than those with no interventions. Dogs progressed quickly through the training schedule (Appendix G) and the differences in behaviour between SD and RP Groups during Sham Dosing further illustrate the positive effects of the Refined protocol on welfare.

Although behaviour for both SD and RP Groups changed between the first and second sham doses, RP Group spent more time interacting with the handler and with the environment, sitting relaxed and tail wagging. SD Group were spending more time 'freezing' or 'collapsed' and with low posture. The aims of the training protocol had been to desensitise the dogs to various aspects of dosing such as the handler, technician, equipment and restraint, and also to associate the procedure pod with positive experiences and rewards. Although some, if not all, dogs exhibited negative reactions to sham dosing, the differences between RP and SD Groups demonstrated a positive effect of the Refined protocol in the RP Group and that desensitisation had occurred.

The differences in behaviour between Groups were not as pronounced during the SD Phase, however there were differences apparent. There was one difference between Control and SD Groups, with Control Group exhibiting more paw lifts, and there were a number of differences with both Groups and RP Group. RP Group dogs spent more time engaging in amicable behaviour and play behaviour, as well less time with half-low posture and sitting alert. As RP Group's behaviour continued to show greater welfare than not only SD Group but also Control Group which had not undergone sham dosing, it can be concluded that the training protocol had not only desensitised dogs to the sham dosing procedure but that it had improved their welfare above that of a Group which had undergone no training or sham dosing, although the differences were less pronounced than during the Training Phase.

During the Dosing Phase, the differences in behaviour between Control and SD Groups became more evident. In the home pen, Control and SD Groups were spending more time with high and half-low posture, and less time with neutral posture and more with high or half-low posture than RP Group. SD Group spent less time resting head down, while Control Group spent less time high tail wagging and exhibited more behavioural events. Both Groups were exhibiting more negative welfare indicators than RP Group, although it appears that their responses were different. RP Group spent less time sitting alert, at the front and with high posture, and more time resting head up or down and with neutral posture. Meanwhile, they also spent less time sitting alert, at the front, low tail wagging, with half-low or low posture and exhibiting fewer behavioural events than and more time tail wagging high and with neutral posture. These are similar differences in behaviour as seen during the Training Phase which suggests that the training protocol

prevented dosing having such a negative effect on RP Group. SD Group exhibit more low tail wagging and low posture than other Groups.

The pattern in behaviour during dosing was similar, with similar differences between Control and SD Groups and RP Group. RP Group spent less time sitting but resisting, struggling or 'freezing' but more time interacting with the handler and sitting relaxed. They also spent less time with high or low posture, or crouching and trembling, and more time with neutral posture. This is a similar pattern as seen during sham dosing. There were a number of differences in key behaviours between Control and SD Groups which suggest that dosing had more of a negative welfare impact on SD Group. SD Group made more escape attempts than Control Group and spent more time crouching. Although not reaching significance, it can be seen in Figures 7.13 and 7.14 that SD Group showed trends towards spending more time 'freezing', less time sitting relaxed, more time with high and low posture, more time trembling and less time with neutral posture. This suggests that the previous exposure to dosing protocol during sham dosing had not habituated SD Group to the procedure, but had rather primed an aversive response due to the lack of control and predictability surrounding sham dosing. In contrast, RP Group had undergone sham dosing but with added aspects of control and predictability, and desensitisation rather than habituation, and this resulted in fewer negative changes in welfare compared to both Control and SD Groups.

7.4.2 Differences in behaviour before and after training and dosing and between Phases

Measuring behaviour before and after training or dosing session did not provide many differences in behaviour, suggesting that responses were stable over time. Similarly, there were few differences seen across Phases, compared to the number of between-Groups differences. Resting head down increased and play decreased after a session, which may be a response to increased activity during a session. Control Group spent less time at the back and resting head down which is likely to be the result of having no sessions in Training and the SD Phases, while SD and RP Groups did. RP Group did not show the increase in high and decrease in neutral posture as did Control and SD Groups.

When looking at the effects of Phase, play and neutral posture decreased during the Dosing Phase, while interacting with the environment increased. Once again, RP Group did not show the changes in high and neutral posture that the other Groups did during the Dosing Phase. Together with the lack of difference between before and after a

session, this suggests that sessions may have had less of an effect on RP Group than on the other Groups.

7.4.3 Welfare Monitoring Tool

The WMT agreed with other measures in a number of ways. It was sensitive to the changes in behaviour which occurred in the Dosing Phase. The home pen score increased on the first day of dosing and decreased until the third day of dosing. During the Dosing Phase, both the dosing score and combined score decrease across the five doses, which agrees with the VAS scores, showing that dogs became increasingly habituated to the procedure across the week. When looking at the combined scores for all Groups over time in Figure 7.18, it can be seen that there appears to be a trend towards SD Group having the highest scores and RP Group having the lowest scores, with Control Group falling between these. This agrees with other behavioural measures which shows SD Group finding dosing more aversive than the other Groups and RP Group having the least negative response to dosing. RP Group had significantly lower scores overall than SD Group, with the difference being marginally non-significant with Control Group. The WMT appears to be a sensitive method of detecting welfare changes which can be implemented by staff.

7.4.4 Mechanical Pressure Threshold Testing

MPT was previously found to agree with cognitive bias testing in that those with a negative affective state were found to have lower MPTs than those with a positive affective state. Its use was designed to measure changing affective states across the course of the study, with the hypothesis being that those who found dosing to be aversive would show a decrease in MPT due to a change in affective state and that those who found dosing less aversive would show more stability and high MPTs.

MPT dropped after the first day of sham dosing and again on the first day of dosing, unsurprising as these were likely to be perceived as the two most aversive events during the study as the dogs were unlikely to be able to predict these events. Control and SD Groups showed decreases in MPT across time, while RP Group showed no significant changes, suggesting stability in MPT and that dosing had less of an effect on them. This agrees with one of the aims of the study, that the training protocol should desensitise the dogs to dosing protocols and that this should mitigate the negative effects of dosing on welfare.

7.4.5 Visual Analogue Scale Scores

VAS scores agreed with WMT scores in that dogs were scored as being more easy to dose across the the Dosing Phase, which agrees with fewer behavioural indicators of negative welfare being scored. As expected, scores increased across the five days of dosing, reflecting some improvement in behaviour across successive doses. However as the scores did not reflect the between-Groups differences in behaviour seen in other measures, it is apparent that those behaviours were not one of the factors influencing the technician's score.

However, VAS scores did not show the between-Group differences that other measures have shown and the technician reported that 'ease of dosing' was influenced by factors other than behaviour. This suggests that as a measure of welfare, VAS scores are not the most suitable to use and that more sensitive measures relating to specific behaviours should be employed.

7.4.6 Time to dose

It was expected that RP Group would be the quickest Group to dose due to the Refinements increasing cooperation with the handler and technician, as McKinley et al. (2003) found when training common marmosets to co-operate with procedures. While this did not prove to be the case, RP were faster to dose than SD, as were Control. This suggests that some factor affecting SD Group increased the time to dose them. Due to the increase in time spent 'freezing' while being dosed, and a clearly observable tension in the jaw while being dosed in several of the SD dogs, it seems likely that this is the reason for the difference. Tension makes it difficult to open the mouth or insert the gavage tube. While the differences in time between Control and RP are subtle (1:09 and 1:08 respectively) and SD (1:17), it did reach significance and should be considered as a factor when weighing up the benefits of a Refined protocol. A slight increase in time to dose per dog as a result of 'freezing' behaviour is not desirable and further supports the conclusion that a Refined or non-sham dosing protocol is of greater benefit than a sham dosing protocol.

7.4.7 Quality of data output

While behaviour shows a clear pattern of results between Groups, the quality of data obtained within Groups must also be high in safety assessment. Food consumption was higher for Control Group in Training and the SD Phases, however the average food consumption for each Group in each Phase did not change. No clear pattern of food

consumption was discernible and so it is concluded that food consumption is robust to changing welfare. This agrees with the lack of discernible effect in food consumption for dog in Chapter 5.

7.4.8 Conclusions

The data presented here suggest that the Refinements to oral gavage had a positive effect on welfare. Dogs in RP Group did not prove to be easier to dose based on the technician's scores, and spent more time interacting with the handler and environment and less time freezing which may be undesirable for an technician. Behaviour in the home pen showed that Refined protocol dosing had a lesser impact on welfare overall when compared to the other Groups. Many of the positive changes seen during the Training Phase were maintained through the Dosing Phase, and there was also a lack of change in MPT, suggesting a lesser impact of dosing on affective state, or at least sensitivity to mechanical pressure, which seemed to increase in the other Groups following dosing. The technician's WMT appears to be a useful way of monitoring welfare for staff, although it requires some further work to achieve agreement with other measures. The benefits of hourly monitoring of the dogs were reported to include increased familiarisation with technician presence and opportunity to observe behaviour without dogs reacting to the technician's presence. The ability to closely monitor individual changes in behaviour is crucial to picking up subtle side-effects in toxicology testing and the use of the WMT encourages identification of individual dogs and their normal behaviour.

The data presented in Chapters 4-6 resulting in the Welfare Assessment Framework suggest that dogs which are less susceptible to changes in welfare following aversive events provide higher quality cardiovascular data. It is therefore recommended that a Refined protocol for dosing by oral gavage like the one described in this study be followed to maximise welfare and data quality. It is also recommended that if it is not possible to provide an adequate pre-study training protocol that sham dosing not be substituted in its place. The WMT proved to be a useful tool for staff to monitor welfare and also allowed staff to become more familiar with the dogs.

CHAPTER 8

General discussion and recommendations

"All difficult things have their origin in that which is easy, and great things in that which is small"

Lao Tzu

While the drive to create new medicines to combat human ill-health demands the use of non-human animals to determine efficacy and safety assessment before use in humans, animals will continue to be used in scientific research. Until we meet Medawar's challenge of full replacement (Stephens, 2011), we have an ethical obligation to minimise harms to the animals used and to obtain the best possible results from animals used in the pursuit of human health. In order to maximise the benefits to humans, and provide an ethical justification, it is essential that the use of animals is Refined wherever it cannot be Replaced and Reduced, so that the best possible value is gained from their use.

In the case of dogs, despite their wide-spread use as a second species in safety assessment testing, there has been little published evidence of the benefits of Refinements in the last decade, despite investment being made in housing and training across industry (Prescott et al., 2004). Without evidence to support benefits to animal welfare and scientific output, the uptake of Refinements will continue to be limited and the benefits to quality of data output questioned. Changes to REACH regulations in the European Union mean that the use of dogs is set to increase in coming years, as well as the potential for increase due to requirements for juvenile toxicity data

(Pellegatti, 2013). The paucity of data on the impact of welfare on data quality when dog use not only exceeds 100,000 dogs per year (see Chapter 1), but is set to increase, means that opportunities for Reduction and Refinement will be missed, and the financial costs associated with loss of sensitivity and increased animal numbers will be amplified.

As a non-prey species, domesticated over thousands of years and selectively-bred to cooperate with humans, dogs should be well adapted to live in an environment which necessitates close contact and cooperation with humans, in contrast to other laboratory-housed species. However, the stressors associated with laboratory housing, husbandry and regulated procedures highlighted in this thesis, even in a Refined environment by global industry standards, suggests that further investment in Refinement is needed. The facility in which this project was conducted had made significant investments in housing and staff training, so the results of the project should be interpreted in this context. It is reasonable to assume that welfare in facilities which have not invested in dog care, particularly in countries where minimum standards are considerably lower (see Chapter 1), would be markedly different from the welfare of these dogs.

There were three aims for this thesis:

(a) Identify reliable indicators of welfare using established measures of affect and behaviour.

(b) Develop a Framework to monitor welfare and data quality, in particular quality of cardiovascular data.

(c) Identify easily-observable measures of welfare which reliably indicate changes in quality of data output and use these to measure the impact of planned Refinements.

The aims of the thesis have been successfully achieved: behaviours occurring in dogs with either a positive or negative affective state were identified at baseline in the home pen, and those most sensitive to changing welfare were identified in response to positive and negative behavioural challenges. Associated patterns of cardiovascular response were identified, including higher blood pressure in the home pen for dogs with negative welfare, and a greater response to some challenges. The effect of the behavioural challenges upon cardiovascular data also provided information as to the effects of common events on data quality, in particular the effect of a brief restraint. The information contained in the Framework was then used to create a technicians' Welfare Monitoring Tool, which was used in conjunction with the Framework to monitor the effects of planned Refinements to oral gavage. The technicians' Tool detected changes

in welfare, as also detected by the Framework, and demonstrated that the Refinements implemented had mitigated some of the negative impacts of gavage on welfare.

8.1 Assessing emotion

Two measures were used to investigate affective state and the relationship with welfare, cognitive bias testing and mechanical pressure threshold testing. It is not possible to have good welfare while also having a negative emotional state (Dawkins, 2008a) and since behaviour may not give a clear indication of emotional state, it was important to evaluate this as a first step. While cognitive bias testing is a fairly recent paradigm in the assessment of non-human animal emotion (see Harding et al., 2004) it is well-established in human psychology. The paradigm has consistently detected changes in affective state in many species (see Chapter 4). This cross-species sensitivity is perhaps due to its ability to make use of the approach/withdraw response of interpretation of ambiguous stimuli and the involvement of emotion in decision making. Although previously utilised in the dog, cognitive bias testing does not present a convenient method of welfare monitoring, particularly in the busy laboratory environment. In this project, the testing of three small populations of dogs took many weeks and required a full-time experimenter to conduct testing. The results however were illuminating and proved to be a central measure in the Welfare Assessment Framework, confirming what was hypothesised: affective state is influenced by a combination of time spent on the unit, housing and husbandry practices, and history of regulated procedures. Although the precise influence of each of these factors could not be determined, the need to determine this and how each influences quality of data output is clear.

Furthermore, mechanical pressure threshold (MPT) showed a pattern consistent with cognitive bias testing, with NAS dogs being more sensitive to mechanical pressure then PAS dogs, as has been shown in the literature on nociception (e.g. Villemure & Bushnell, 2002, see Chapter 4). MPT was also found to change over time as the result of administering doses by oral gavage, although this change was dependent on whether or not Refinements had been made to the pre-study protocol (see Chapter 7). This is in contrast to findings in a population of beagles tested over six months by Hoffmann et al. (2010), where the threshold was found to be stable. It therefore seems likely that MPT is stable over time, except where events lead to a change in affect; in Chapter 4, Stock dogs had a higher mean threshold than the longer-term groups. Although MPT was not recorded beyond the end of the study in Chapter 7, observation of their behaviour by staff over the following six months suggested that the decrease in welfare

observed during the study was not long-lasting, so it may be that MPT recovered after the study.

Given that it appears to be possible to mitigate this increased sensitivity in the case of oral gavage, by implementing simple Refinements, it is strongly recommended that such Refinements be made to dosing protocols, and investigated in other regulated procedures. An impact of such increased sensitivity on the battery of measures used in safety assessment testing is also likely to be present, although it was outwith the scope of this project to investigate this fully due to lack of access to animals on-study, or to conduct blood sampling.

It is not clear if affective state is the cause or result of the increased sensitivity, but what is clear is that in the SP dog group which had undergone the most intensive and prolonged use, affective state was more negative than in other Groups and dogs were considerably more neophobic than other dogs, particularly in response to unfamiliar people or equipment. These dogs showed differences in baseline cardiovascular function which varied by Affective State and also showed undesirable changes in cardiovascular function in response to behavioural challenges (see Chapter 6). This suggests that dogs with a NAS are unlikely to produce consistently good quality data, especially where both PAS and NAS dogs are present in a study cohort.

8.2 Investigating welfare in the home pen and in response to challenges

The investigation of welfare in the home pen confirmed the differences in welfare found by the investigation of Affective State and also provided behavioural parameters which can be used to assess the welfare of dogs. Behaviours which are known to vary with welfare state in humans and other nonhuman animals, such as vigilance (alert, Rushen et al., 1999), agitation (persistent high posture, moving between pens, Hubrecht et al., 1992) and behavioural events such (paw lifts, lip smacking, Beerda, Schilder, Van Hooff et al., 1999) were found to occur more often in dogs with negative affective states and following negative events. Conversely, behaviours such as restfulness (resting head up or down), positive interactions with conspecifics or interacting with the environment were found more often in dogs with positive affective states (Boissy et al., 2007).

It is not surprising given the differences in baseline behaviour that baseline blood pressure was elevated in NAS dogs. Behaviours indicating negative welfare were more prevalent and NAS dogs also tended to be more active (standing rather than resting). Analysis of 95% CIs showed that there was little overlap in CIs when comparing PAS

and NAS dogs, and there was also a medium-sized effect size (0.60 and 0.66) for SBP and DBP, showing that welfare has had a detectable effect on blood pressure. The effect on HR was less pronounced, with PAS dogs tending to have lower heart rates, producing a small effect size (0.39). These data suggests that welfare state must be factored into any analysis of these parameters, otherwise there is a risk of drawing incorrect conclusions. The potential health impact of increased blood pressure must also be considered, as long-term increased blood pressure in repeated-use dogs may result in reduced health, changing the response of organ systems to the test compound.

Measuring of affective state and baseline welfare suggested that differences in life history had influenced welfare between the three groups of dogs, conducting four challenges clarified this further. Stock dogs, having experienced no regulated procedures or long-term housing on the unit showed few signs of negative welfare in response to any of the challenges, and showed signs of positive welfare in response to the two positive challenges. SP dogs showed greater levels of excitement in response to human interaction, while DMPK dogs showed signs of negative welfare in response to a brief single-housing, reflecting different experiences of each of these events (see Chapter 3). Brief restraint elicited several signs of negative welfare in SP and DMPK dogs, some of which lasted for up to 30 minutes after the event. Many of the indicators of negative welfare which emerged in response to the challenges were more pronounced or only present in NAS dogs, which is consistent with reduced welfare, or the ability to cope. The differences which emerge when comparing PAS and NAS dogs can be attributed to a greater ability to cope with various events in the PAS dogs compared to the NAS dogs. Predictability and control are known to be important for promoting positive welfare (Bassett & Buchanan-Smith, 2007) and given the higher prevalence of PAS dogs in the Stock dog group, it seems likely that experience of uncontrollable unpleasant events in the SP and DMPK dogs' lives has reduced their ability to cope.

8.3 Refining oral gavage

Having identified measures of welfare at baseline in the home pen, and identifying those measures sensitive to changing welfare through the employment of challenges it was important to demonstrate that these can be used practically to monitor welfare, and also to measure the impact of Refinements. Responses to human interaction and restraint suggested that these were both areas in need of Refinement and which had to potential to improve welfare. The effect of welfare on cardiovascular parameters also highlighted the need to improve welfare to prevent the patterns of negative welfare seen in the long-term dogs in SP and DMPK groups.

Oral dosing is the most common route of dosing in the dog and was experienced many times by the dogs with histories of regulated procedures in the Framework studies. It is therefore reasonable to assume that it was a factor in the poor welfare exhibited by some dogs. The literature on dog training, resilience and welfare suggests that harmonisation of welfare in dogs should be achievable through Refinements relating to human interaction, positive reinforcement training and predictability. The findings of this study (Chapter 7) showed that sham dosing does not desensitise dogs to oral dosing (and may prime a more negative response), and that a Refined protocol can prevent the emergence many of the negative welfare indicators seen in dogs subject to regulated procedures (Chapter 5). Perhaps most importantly, the methodology employed for both training and welfare monitoring by the responsible technician were practical to employ, providing considerable benefit for the level of investment required. The findings from this study prompted the Industrial Partner to implement a training and welfare monitoring protocol using this protocol for all dogs being held, which also resulted in significant improvements in welfare in response to handling over a four-week period (Appendix G).

8.4 Recommendations

8.4.1 Enrichments to improve welfare

The responses of the dogs to the positive challenges in Chapter 6, human interaction and the food toy, suggest that these are effective Refinements for improving welfare. Enriching dogs' environments with toys, particularly those which encourage chewing or foraging, has been recommended in the pet dog literature (Wells, 2004a) as well as for laboratory-housed dogs (Hubrecht, 1993; Prescott et al., 2004). This also represents a practical method of improving welfare which requires little staff input: the toy need only be filled and replaced once a day during normal husbandry duties.

Positive human interaction has been shown to have numerous benefits for dog welfare and health (see Chapter 2 for review), however the implementation in laboratory environments has been varied and sometimes limited due to perceived lack of benefits and frequent publication of recommendations (e.g. Prescott et al., 2004). There may also be concerns about staff presence adversely affecting the dogs' level of excitement and so staff contact may be reduced. The human interaction challenge demonstrated this effect, with all of the dogs showing signs of positive welfare in response to the HI period, but with some dogs becoming greatly over excited by it (high posture, rapid movements, climbing on the experimenter). Rather than being viewed as a variable to

be controlled, human interaction should be used to both improve welfare and to acclimatise the dogs to staff presence. During data collection for Chapter 7, the technician noted that his frequent presence in the unit due to collecting data hourly resulted in the dogs no longer becoming excited by his presence. The high heart rates exhibited by some SP dogs during the HI period suggests that more frequent contact would be beneficial for reducing excitement and may reduce the need to prevent staff presence during data collection. Stock dogs had the best welfare, yet experienced little additional human contact. This suggests that positive human interaction is increasingly important for dogs undergoing regulated procedures and held on the unit for long periods of time.

8.4.2 Refinements to improve welfare

While the challenges illustrated the potential benefits of Refinements to welfare, areas in need of Refinement were also found. The different patterns of response of dogs to single-housing showed that the experiences associated with previous single-housing had an influence on welfare. While SP dogs showed no change in behaviour, their welfare was considered generally more negative than other dogs, which may be in part attributable to frequent periods of single-housing. DMPK dogs showed a more negative response, which was unsurprising since for them, restriction to single-housing was associated with regulated procedures. Stock dogs showed little response, suggesting that short periods of single-housing alone are not aversive. Refinements to the acclimatisation protocols for single-housing, by desensitising the dogs to the experience, may mitigate these changes in welfare seen in the longer-term colonies.

The area most in need of Refinement, however, was restraint. While all dogs were habituated to restraint using a standard Industry protocol, those which had experienced restraint associated with regulated procedures (SP and DMPK groups) showed signs of negative welfare not otherwise seen (trembling, crouching, urinating), in particular those dogs with negative affective states. The high heart rate and blood pressure values achieved by some SP dogs during restraint were of concern both from an animal welfare perspective and because of potential interactions with a test compound where cardiac effects are of interest. The blood pressures achieved were also capable of causing target organ damage (Brown et al., 2006), although without pathology end points (no dog was on a terminal study), it is not possible to quantify this. These findings prompted the Refinement of oral gavage explored in Chapter 7.

8.4.3 Positive reinforcement training

The findings in Chapter 7 strongly support the implementation of a positive reinforcement training (PRT) protocol for desensitisation to restraint. The protocol employed in Chapter 7 (four sessions of PRT), resulted in dogs which would sit calmly on the table with little physical restraint, a protocol which could easily be implemented for other regulated procedures and for health check and other husbandry. The welfare of the dogs in the RP group continued to improve across the course of the study as a result of the increased positive interaction with staff, mitigating the negative impact of dosing. AstraZeneca has fully integrated the protocol into standard operating procedures for stock dogs as a result of the findings, with care staff achieving similar results in four training sessions (see Appendix H). It is recommended that a similar protocol is adapted for stock dogs within the pharmaceutical and chemical industries, in particular at an early age, to ensure that welfare is not negatively influenced by aversive events.

8.4.4 Theoretical recommendations

One of the clearest theoretical recommendations which arises from this project is the need for integrative, multi-factorial assessment of welfare. Simply comparing the behaviour of the three groups (SP, DMPK and Stock) at baseline would have elicited little information about welfare state, rather reflecting the different housing and husbandry practices of the groups. The integration of cognitive bias testing and later mechanical pressure threshold testing provided information about affective state, and perhaps more generally an ability to cope with the environment. When the two affective states (positive and negative) were compared, a pattern of welfare emerged, along with associated behavioural indicators and patterns of cardiovascular function. Monitoring these in response to challenges helped to identify the most sensitive measures and provided information which was used to develop Refinements.

The Welfare Monitoring Tool proved to be successful in monitoring the effects of Refinements to oral gavage (Chapter 7) and agreed with changes in welfare detected by the Framework. However, this is just one of the events in the life cycle of the laboratory-housed dog with the potential to reduce welfare. A number of other very common events (including the human interaction, feeding toy and single-housing challenges in Chapter 6) remain to be implemented fully in the working laboratory environment at a facility-wide level. Other common potentially aversive events including transport and rearing practices were outwith the remit of this study but remain key events in the lifetime of the laboratory-housed dog. It is strongly recommended that the Welfare

Assessment Framework and Welfare Monitoring Tool are implemented in practice to determine the benefits of Refinements recommended here and elsewhere in the literature.

Furthermore, while a link between welfare and data output was found, particularly in that cardiovascular data varied between the positive and negative welfare types, these data are insufficient to quantify the full impact of welfare on data output. The blood pressures achieved by dogs in response to restraint are capable of causing target organ damage if sustained over time, however without examining pathological endpoints in dogs subjected to repeated regulated procedures, the nature of this impact cannot be assessed. Similarly, while links between enduring stress and immunology have been found, these parameters were not examined in this project. The methodology employed in this project was to use existing activities within the dog facility to examine the impact on welfare and difficulties occurred in obtaining studies suitable to collect these physiological data while also collecting the behavioural and affective data necessary to determine welfare state. It is recommended that the Welfare Assessment Framework is applied to studies in which these data are collected in order to provide information on how welfare interacts with key safety assessment variables under study conditions.

8.5 Final conclusions

One of the strongest outputs of this thesis is that it serves to highlight how little we know about the welfare of dogs in laboratories. The paucity of empirical data on which to base recommendations is apparent and yet ample studies suggesting that there should be a link between welfare and quality of data output were found (Chapter 2). With the exception of the many studies carried out in the 1990s by Robert Hubrecht and Bonne Beerda and colleagues, we know little about how dogs perceive the laboratory environment. Additionally, many recommendations based upon expert opinion and company practice have been made, for example the extensive Joint Working Group on Refinement report (Prescott et al., 2004) without follow-up studies to provide evidence of their efficacy.

The Industrial Partner in this project had made many modifications to its facility and practices based on the available best practice guidelines. At the time of its design, the unit was the first to employ many of the improvements to housing highlighted in Section 3.4. Dogs were also bred on-site, removing many stressors associated with transfer between facilities and acclimatisation. Further, at the time of data collection for this project, the use of animals in scientific procedures was governed in Europe by Directive 86/609/EEC, and in the UK by A(SP)A (1986). This means that the

findings presented here were conducted under regulations which were largely of a higher standard than other countries using dogs in Europe (now governed by Directive 2010/63/EU). Of the more than 100,000 dogs used globally, only around 3% are used in the UK under these standards and around 20% in Europe under the new standards.

The detection of both positive and negative welfare states in this population of dogs housed in what should be considered very good conditions should be concerning alone from an ethical standpoint. However effects of welfare on quality of data output were also detected. Differences in baseline blood pressure and mechanical pressure threshold were found in PAS and NAS dogs. This means that welfare is acting as a confound, despite the dogs being raised in near-identical conditions and experiencing similar life events. Dogs with negative welfare were also more susceptible to changes in behaviour and cardiovascular parameters when presented with positive and negative events. This is likely to reduce the sensitivity of measures. The magnitude of the effect of welfare at baseline and in response to the restraint challenge in particular means that welfare is a factor which cannot be discounted in ensuring 'good science' is conducted.

It is reasonable to assume that if these differences were found in this population of dogs, subject to fewer stressors than the majority of laboratory-housed dogs used globally, that much greater differences would be found between dogs with positive and negative welfare. Negative welfare is also likely to be more negatively-valenced in global dog populations and therefore the effects of welfare on quality of data output are likely to be greater. This should be particularly concerning for the reproducibility and repeatability of data, as well as the sensitivity where ceiling effects occur. Although not investigated in this project, the effect of welfare on other parameters in the safety assessment battery should be investigated. Welfare is known to influence immune function (Everds et al., 2013), while hypertension is known to have an impact on target organ function (Brown et al., 2006), suggesting that these systems are also likely to be impacted. Everds et al. (2013) provides a comprehensive list of the known effects of welfare on parameters of interest, and it is clear that in the dog, we do not sufficiently understand the effects of welfare on these parameters.

However, it was also shown that with simple Refinements such as desensitisation and predictability, many of the undesirable effects of aversive events could be mitigated, preventing a negatively-valenced change in welfare. The protocol employed in Chapter 7 increased the resilience of dogs to oral gavage. It is apparent from examining dogs in SP and DMPK groups that identical treatment can lead to different outcomes because individuals may be more or less able to deal with potential stressors. Comparison with dogs receiving different treatments and those receiving the Refined protocol suggests that the Refinements resulted in harmonised welfare. The literature suggests that

simple interventions can make considerable differences to dog welfare (e.g. Prescott et al., 2004; Laule, 2010) and this proved to be the case in this project.

It is strongly suggested that recommendations for Refinements to the life cycle of the laboratory-housed dog made here and elsewhere (e.g. Prescott et al., 2004) are fully investigated, with the impact on welfare and data output quantified using the Welfare Assessment Framework. The findings of this project suggest that welfare can potentially have a considerable effect on cardiovascular data and this evaluation should be replicated in other parameters of interest. The dog has evolved to be a working animal through our selective breeding, and the implementation of training and desensitisation programmes has the potential to be an efficient and effective method of harmonising welfare between individuals in the laboratory environment.

APPENDIX A

Results of cognitive bias testing

This Appendix provides full analysis of cognitive bias testing results from Chapter 4. All dogs had demonstrated significantly longer latencies to the unrewarded (UR) stimulus box than the rewarded (R) stimulus box during testing, thus confirming that dogs could distinguish between them and that the results are valid. Only one dog (Ringo) demonstrated latencies to the probe boxes which differed significantly, so the analysis focuses on the pattern of responses: "optimistic" and responding quickly to all stimulus boxes, or "pessimistic" and responding slowly to all but the rewarded stimulus box.

A.1 SP Dogs

Bert

Optimist Bert exhibited very short mean latencies to all boxes in testing (<5 sec of the maximum 30 sec), with the only longer latency being to the unrewarded box.

George

Optimist Similarly, George also exhibited very short latencies to all boxes (<5 sec), with only the unrewarded box eliciting a significantly longer latency.

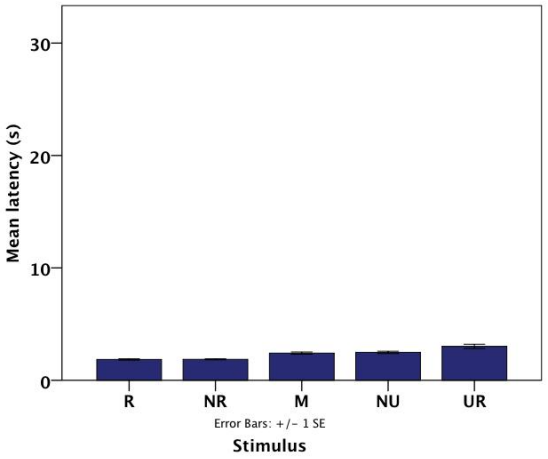


FIGURE A.1: Mean latency to each stimulus box for Bert

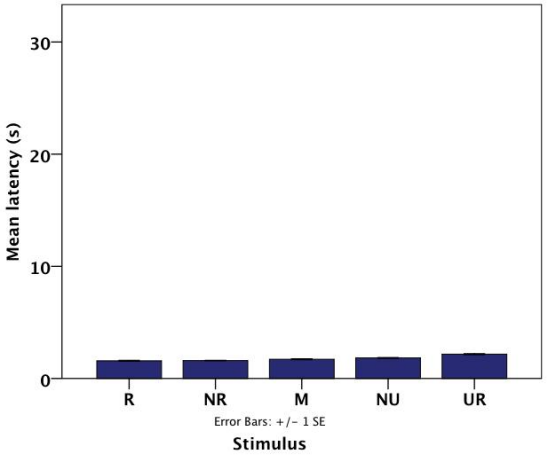


FIGURE A.2: Mean latency to each stimulus box for George

Bob

Optimist Bob exhibits a slightly different pattern of responses, with very short mean latencies to the rewarded, near rewarded and middle boxes (<5 sec), with a slightly longer latency to the near unrewarded box and longer again to the unrewarded box. He demonstrated an optimistic response when the box was rewarded or ambiguous, but this response became more conservative when the box was unrewarded.

Peewee

Optimist Peewee was another dog to exhibit a highly optimistic response, with latencies to all boxes being less than 5 seconds, and only the latency to the unrewarded box being significantly longer.

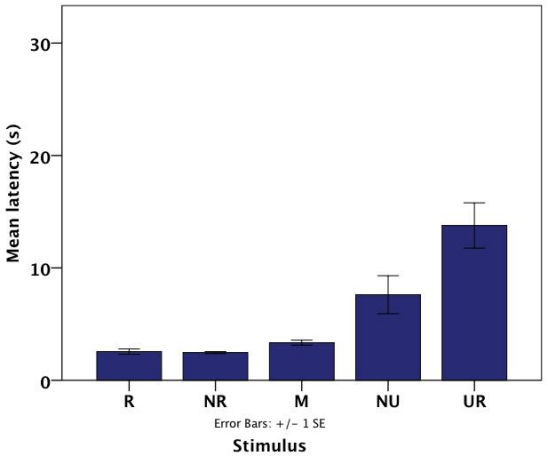


FIGURE A.3: Mean latency to each stimulus box for Bob

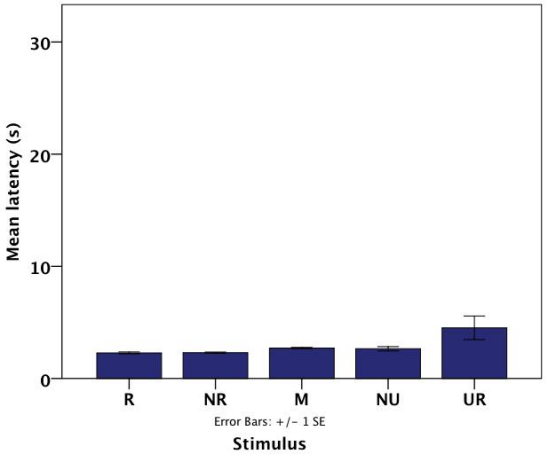


FIGURE A.4: Mean latency to each stimulus box for Peewee

Ringo

Pessimist The difference in latencies exhibited by Ringo and the other SP dogs is clearly demonstrated in Figure A.5. Although he exhibited a very short mean latency to the rewarded box, all other boxes elicited much longer latencies, approaching the maximum latency of 30 seconds which reflects the high number of responses which exceeded this maximum.

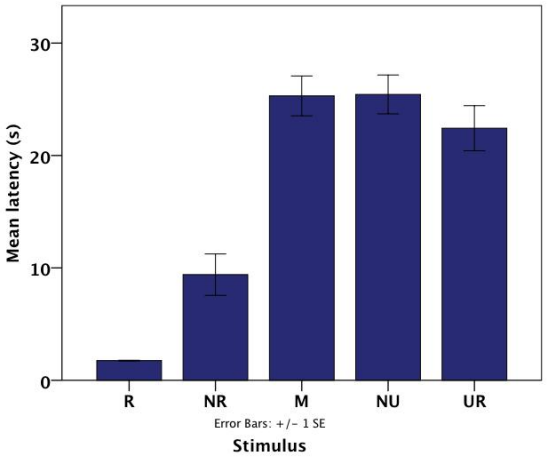


FIGURE A.5: Mean latency to each stimulus box for Ringo

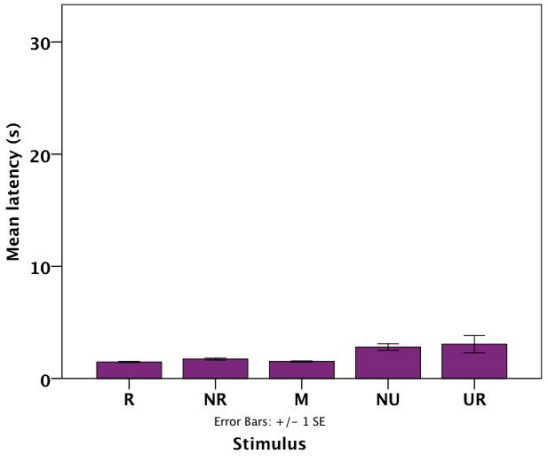


FIGURE A.6: Mean latency to each stimulus box for F25

A.2 DMPK dogs

F25

Optimist F25 exhibited the highly optimistic pattern of responses across all boxes with mean latencies less than 5 seconds.

F26

Optimist F26 exhibited very short mean latencies to the rewarded and near rewarded boxes (< 5 sec), slightly longer latencies to middle and near unrewarded boxes (<10 sec) and a significantly longer mean latency to the unrewarded box (<15 sec).

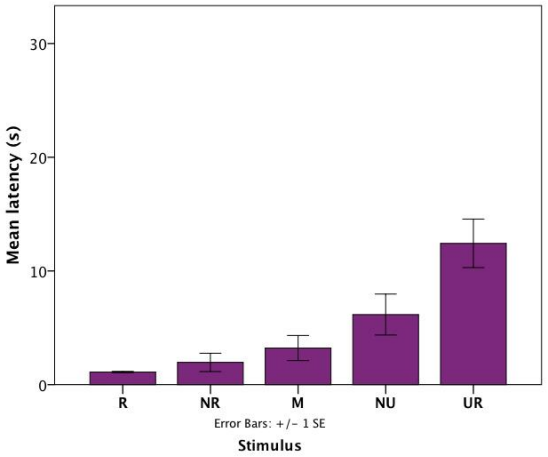


FIGURE A.7: Mean latency to each stimulus box for F26

F37

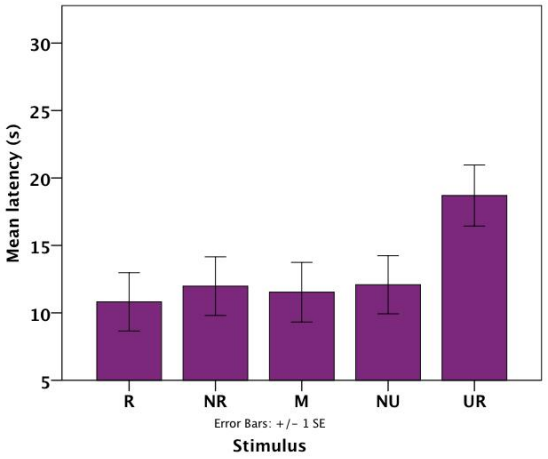


FIGURE A.8: Mean latency to each stimulus box for F37

Optimist F37 exhibited consistent mean latencies across all boxes other than the unrewarded box. Although these latencies were slightly longer than for other dogs ($\sim$ 12 sec), they were significantly shorter than to the unrewarded box ($\sim$ 20 sec).

M12

Optimist M12 exhibited a pattern of responses different from other dogs, in that although his mean latency to the rewarded box was significantly shorter than to the unrewarded box, his responses to the probe boxes were shorter again.

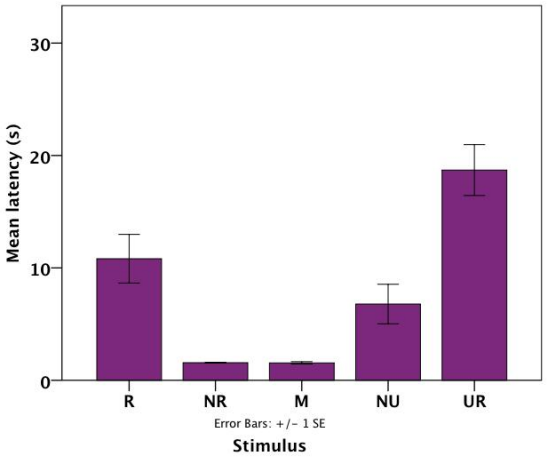


FIGURE A.9: Mean latency to each stimulus box for M12

M13

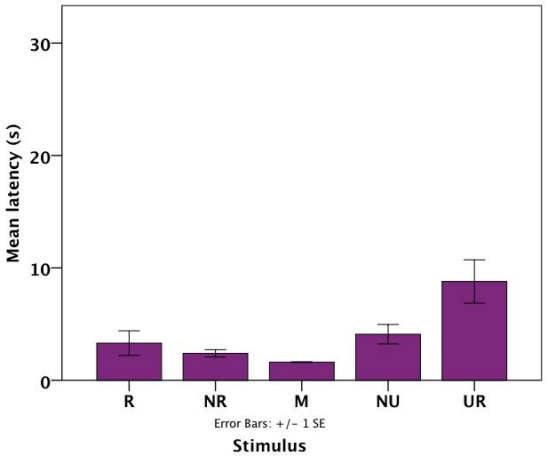


FIGURE A.10: Mean latency to each stimulus box for M13

Optimist M12 also exhibited very short mean latencies (<5 sec) to all but the unrewarded box.

A.3 Stock dogs

F268

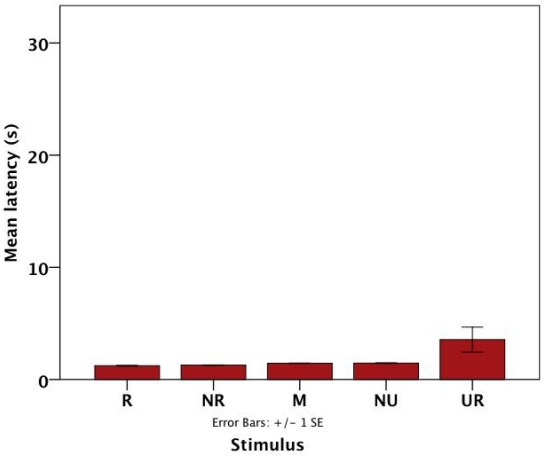


FIGURE A.11: Mean latency to each stimulus box for F268

Optimist F268 also exhibited the highly optimistic pattern of responses, with very short mean latencies (<2 sec) to all but the unrewarded box.

F273

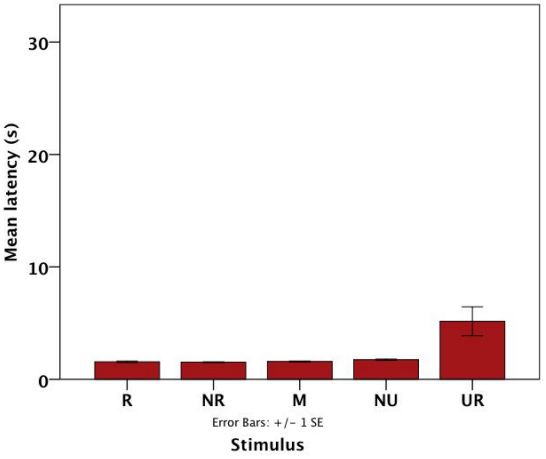


FIGURE A.12: Mean latency to each stimulus box for F273

Optimist F273 exhibited the same highly optimistic pattern of responses, with very short mean latencies (<2 sec) to all but the unrewarded box.

F366

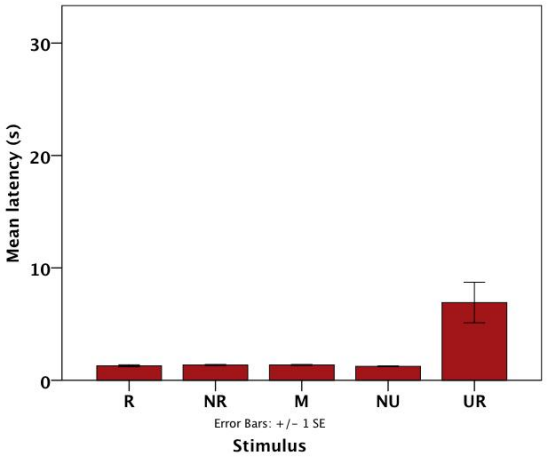


FIGURE A.13: Mean latency to each stimulus box for F366

Optimist F366 also exhibited the same pattern of responses, with very short mean latencies (<2 sec) to all but the unrewarded box.

M206

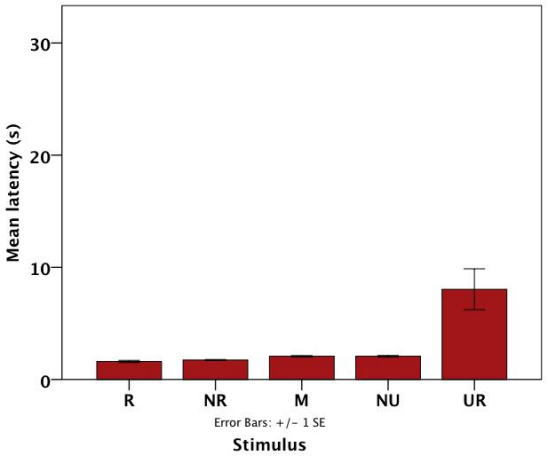


FIGURE A.14: Mean latency to each stimulus box for M206

Optimist M206 also exhibited the same pattern of responses, with very short mean latencies (<2 sec) to all but the unrewarded box.

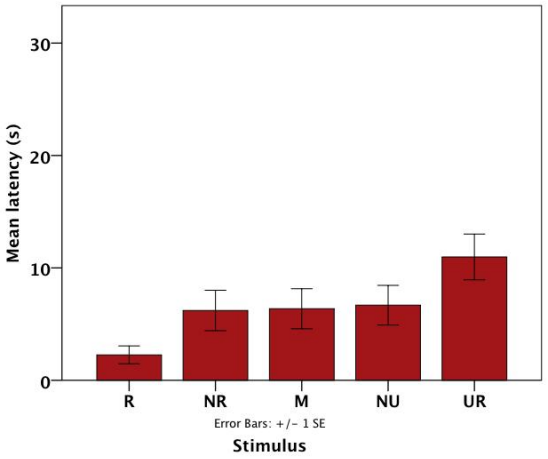


FIGURE A.15: Mean latency to each stimulus box for M292

M292

Pessimist M292 showed a more ambiguous pattern of responses than other dogs. All stock dogs exhibited very similar latencies to all but the unrewarded box, while M292 demonstrated longer latencies to all other boxes when compared to the rewarded box, indicating that they were interpreted less "optimistically". Although these latencies are still short (out of the maximum 30 sec) the latencies to all boxes other than rewarded are clearly more similar to the unrewarded box, in other words he interpreted ambiguous probe boxes as being more likely to be unrewarded than rewarded.

M1

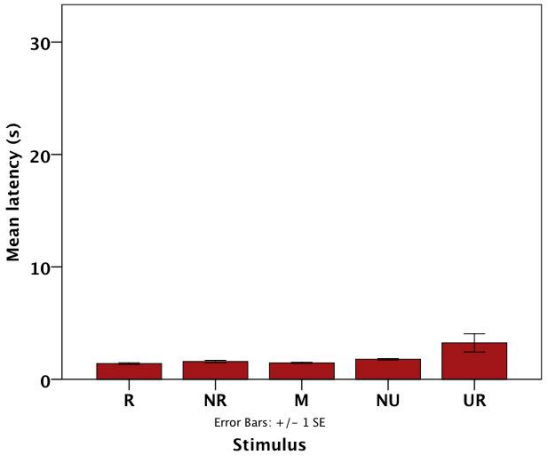


FIGURE A.16: Mean latency to each stimulus box for M1

Optimist M1 exhibited the same pattern of responses seen in other optimistic dogs, with very short mean latencies (<2 sec) to all but the unrewarded box, and a mean latency to the unrewarded box of less than 5 seconds.

APPENDIX B

Behavioural indicators of welfare

TABLE B.1: Locations within the home pen

Location	Description
Front	Within the front half of the pen
Back	Within the rear half of the pen
Barrier	At the hatch in the barrier between two adjoining pens

TABLE B.2: Behavioural measures of positive welfare in the home pen

Behaviour	Description	Source
Resting head up	Sitting or lying, not apparently asleep but not orientated towards any stimulus	1, 6
Resting head down	Lying, may be apparently asleep, not orientated towards any stimulus	1, 2, 4, 6
Interact with environment	Sniffing or investigating pen or objects	1
Amicable	Lick, play, allogroom dog, often with tail wag	2, 7, 8
Solicit play	Bow, metaplay	2, 3, 4
Play (self)	Usually involving toys or other objects	
Play (social)	Bouncing gait, play face, wrestle, play chase	2, 3, 4, 7, 8
Calm locomotion	Walk, 4 Beat gait and 3 feet on the ground at any one time	12

Table B.3: Behavioural measures of negative welfare in the home pen

Behaviour	Description	Source
Autogrooming	Cleans self using mouth and/or paws	1, 6
Flank sucking	Takes skin in mouth and sucks	2, 3, 11
Digging	Using the paws to repetitively dig at a surface or object	1
Destruction of environment	Using teeth or paws to tear or damage pen or objects	1
Stand walls	Stands on hind legs with forelegs against wall	1, 2, 3, 4, 5, 6, 7
Threat	Snarl, raise hackles to dog	2, 3, 4, 8
Attack	Bite, snap, paw or chase dog	2, 3
Defend	Evade dog, cower, roll over, lick face	2, 3
Competitive	Defend object or food from dog	2, 3
Circling	Repetitive movement around pen	1, 2, 3, 5, 6, 7
Pace	Repetitive pacing, usually along a boundary	2, 3, 6
Social pace	Repetitive pacing, in parallel with a dog on other side of boundary	2, 3
Sit alert	Dog orientated towards stimulus while in a sitting position	14
Stand alert	Dog orientated towards stimulus while in a standing position, usually accompanied by high posture	14
Rapid locomotion	Trot, 2 Beat gait, diagonally opposite legs move together	12

Table B.4: Postural measures of welfare in the home pen

Behaviour	Description	Source
High	Breed specific posture as shown under neutral conditions, with the addition of high tail, head and ear position	1, 5
Neutral	Breed specific posture as shown under neutral conditions	1, 12
Half-low	Two features from: low position of tail, backwards bending of ears, bent legs	1, 5, 6
Low	As above, all three features present	1, 5, 6
Very low	As above, with body close to ground	1, 6
Tail wag high	Repetitive movements with the tail held high	1, 5, 6, 9
Tail wag low	Repetitive movements with the tail held low	1, 5, 6, 9

Table B.5: Behavioural events indicating negative welfare in the home pen

Behaviour	Description	Source
Startle	Sudden jump in response to stimulus	7
Body shake	Whole body shivers, trembles	1, 5, 7
Oral behaviours	Includes tongue out, snout licking, swallowing, lip smacking	1, 5, 6
Paw lift	Sudden raising of one limb, usually foreleg, and usually in response to stimulus	1, 5, 6, 7, 10
Yawn	The mouth is opened wide and a long deep breath is taken	1, 5, 6, 7
Pant	Open mouth, rapid breathing with tongue extended	1, 6
Jump	All four limbs leave ground simultaneously	2, 3, 4
Wall bounce	Dog jumps towards wall and contacts with limbs	2, 3
Jerk	Sudden movement, usually away from a stimulus	7
Circle	Singular rapid movement around pen	7

Table B.6: Other behavioural measures in the home pen

Behaviour	Description	Source
T-dog	Muzzle placed across neck of another dog	2, 3
Crouch	Bent legs, body lowered towards ground	1
Tremble	Clear shivering of the body	1
Vocalise	Barking, growling, soft whining, loud whining, low pitched or high pitched vocalisations, yelping.	1, 6

Table B.7: Additional behavioural measures for Challenges

Behaviour	Description	Source
Human interaction		
Proximity	Dog stands or sits close to experimenter but does not interact	
Petting	Dog allows stroking and grooming by experimenter	
Petting & standing	Dog stands on experimenter while being petted	
Avoid proximity	Dog actively avoids contact with experimenter	
Location		
Other pen	Dog is not in the same pen as experimenter	
Same pen (proximate)	Dog is in the same pen and is close to the experimenter	
Same pen (distal)	Dog is in the same pen but at a distance	
Threat human	Snarl, raise hackles to human	2, 3
Attack human	Bite, snap, paw or chase human	2, 3
Defend human	Evade human, cower, roll over, lick face	2, 3
Competitive human	Defend object or food from human	2, 3
Amicable	Lick, play, allogroom human, often with tail wag	13
Solicit play human	Bow, metaplay	9
Play human	Usually involving toys or other objects	13
Feeding toy		
Play with FT	Dog interacts with FT as opposed to other toys in the pen	
Carry FT	Dog takes the FT in mouth and moves around with it	
Additional location	Modifiers 'same pen' and 'other pen' added to existing locations	
Restraint		
Avoid/struggle	Dog attempts to avoid restraint and/or human	13
Location		
On table	Dog on table/being held by experimenter	
On floor	Dog on floor having left table	
Single-housing		
Proximity seeking	Dog attempts to contact conspecific through pen bars	

1. Beerda, B., Schilder, M., Van Hoo, J., De Vries, H. Mol, J. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, 58 (3-4), 365-381.

2. Hubrecht, R. (1993). A comparison of social and environmental enrichment methods for laboratory housed dogs. *Applied Animal Behaviour Science*, 37 (4), 345361.

3. Hubrecht, R., Serpell, J. Poole, T. (1992). Correlates of pen size and housing conditions on the behaviour of kennelled dogs. *Applied Animal Behaviour Science*, 34 (4), 365383.

4. Spangenberg, E. M. F., Bjorklund, L., Dahlborn, K. (2006). Outdoor housing of laboratory dogs: Effects on activity, behaviour and physiology. *Applied Animal Behaviour Science*, 98(3), 260-276.

5. Beerda, B., Schilder, M., Van Hooff, J., De Vries, H. Mol, J. (1999). Chronic stress in dogs subjected to social and spatial restriction. I. Behavioral responses. *Physiology and Behavior*, 66 (2), 233-242.

6. Haverbeke, A., Laporte, B., Depiereux, E., Giffroy, J.-M. Diederich, C. (2008). Training methods of military dog handlers and their effects on the team's performances. *Applied Animal Behaviour Science*, 113 (1), 110-122.

7. Schipper, L. L., Vinke, C. M., Schilder, M. B. Spruijt, B. M. (2008). The effect of feeding enrichment toys on the behaviour of kennelled dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, 114 (1), 182-195.

8. Horowitz, A. C. (2002). *The behaviors of theories of mind, and a case study of dogs at play* (Unpublished doctoral dissertation). University of California, San Diego.

9. Normando, S., Corain, L., Salvadoretti, M., Meers, L. Valsecchi, P. (2009). Effects of an enhanced human interaction program on shelter dogs' behaviour analysed using a novel nonparametric test. *Applied Animal Behaviour Science,* 116 (2-4), 211-219.

10. Stephen, J. Ledger, R. (2005). An audit of behavioral indicators of poor welfare in kennelled dogs in the United Kingdom. *Journal of Applied Animal Welfare Science*, 8 (2), 79-96.

11. Beerda, B., Schilder, M., Van Hooff, J. De Vries, H. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science* , 52 (3-4), 307-319.

12. Overall, K. L. (2014). The ethogram project. *Journal of Veterinary Behavior: Clinical Applications and Research*, 9 (1), 1-5.

13. McGreevy, P. D., Starling, M., Branson, N., Cobb, M. L. Calnon, D. (2012). An overview of the doghuman dyad and ethograms within it. *Journal of Veterinary Behavior: Clinical Applications and Research*, 7 (2), 103-117.

14. Ley, J., Bennett, P. Coleman, G. (2008). Personality dimensions that emerge in companion canines. *Applied Animal Behaviour Science*, 110 (3), 305-317.

APPENDIX C

24-hour pilot data

A pilot study was conducted to determine both the best time of day to conduct behavioural observations, and the most suitable behavioural measures to use. Data were collected over a 24-hour period on six dogs housed in groups of two or three in the home pen. Five of these six dogs were the same as the SP dogs used in the studies in Chapters 4-6. Dog 660 which took part in this study was subsequently replaced by Dog 497 (Nibbler) before the collection of data for the Welfare Assessment Framework, while dogs 200 (Bouncer) and 207 (Ringo) were added later.

A behavioural coding scheme similar to that presented in Appendix 1 was used. Some behaviours not previously defined in the coding scheme were noted as occurring frequently and were therefore added (see Appendix 1). Heart rate and blood pressure data were recorded from implanted telemetry (for details see Chapters 3 and 5). Cardiovascular data were sampled every fifteen minutes as an average of one minute of continuous data. Individual behaviours were not of as much interest as the overall level of activity, so behaviours were assigned a score based on the associated activity level: 0 (resting head down, no movement); low (sitting, awake but not moving); medium (calm locomotion, interacting with environment, gentle activity); high (rapid locomotion, pacing, play, high-energy activity). Figure C.1 shows activity level plotted across 24 hours, with Figures C.3 and C.2 showing heart rate and blood pressure respectively. Minimum and maximum values are shown (lighter blue lines) as well as the mean (dark blue line). Changes in heart rate and blood pressure reflected changes in activity level, with activity being fairly constant in the morning from around 6am until feeding around 11am. Activity was low during feeding, with a spike in activity associated with return to group housing around 1pm. Activity decreased and remained low throughout the night.

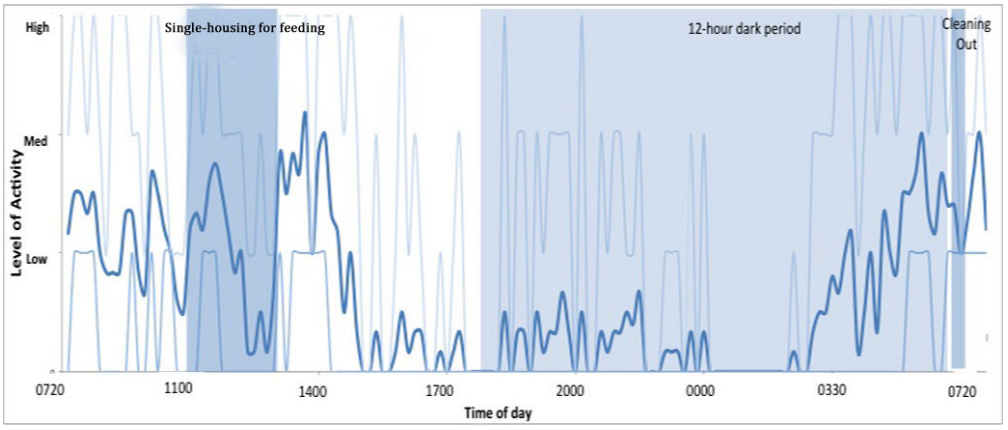


FIGURE C.1: Mean activity, with maximums and minimums over 24 hours

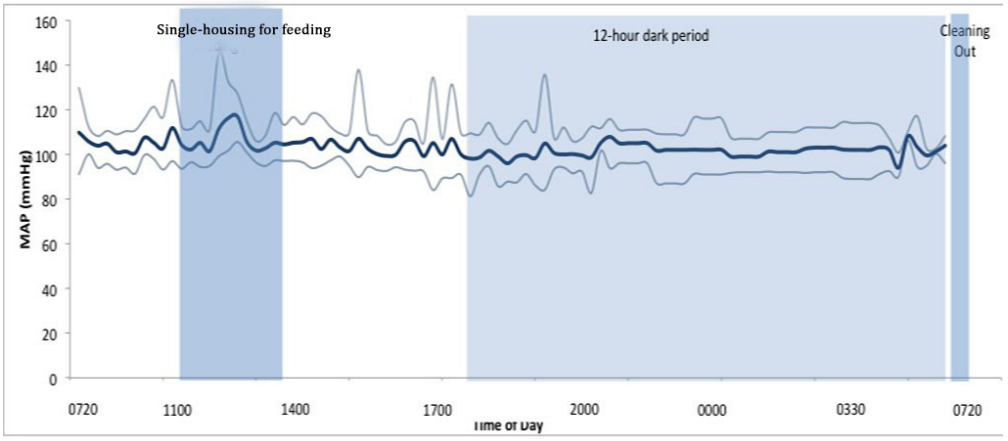


FIGURE C.2: Mean arterial pressure (mmHg), with maximums and minimums over 24 hours

An activity budget for the dogs over 24 hours is shown in Figure C.4. Resting head down accounted for 53% of time. when behaviour was corrected for time resting head down, time awake was predominantly spent alert (38%) with resting head up (13%), interacting with the environment (12%), amicable dog interactions (12%) and play (10%) making up the majority of other behaviours. While play and amicable dog interactions were seen less frequently in baseline behaviour (Chapter 5), this is explained by a peak in activity at approximately 3am when dogs engaged in social interactions and play behaviour.

As a result of these analyses, it was determined that the best time for data collection was between 7.30am and 11.30am, given the consistency in activity during these times. It is worth noting that the fairly simplistic analysis employed here did not detect the differences in welfare later detected, nor the effects of welfare on cardiovascular data. This only serves to highlight the importance in using a multi-faceted approach and ensuring appropriate sensitivity in measures.

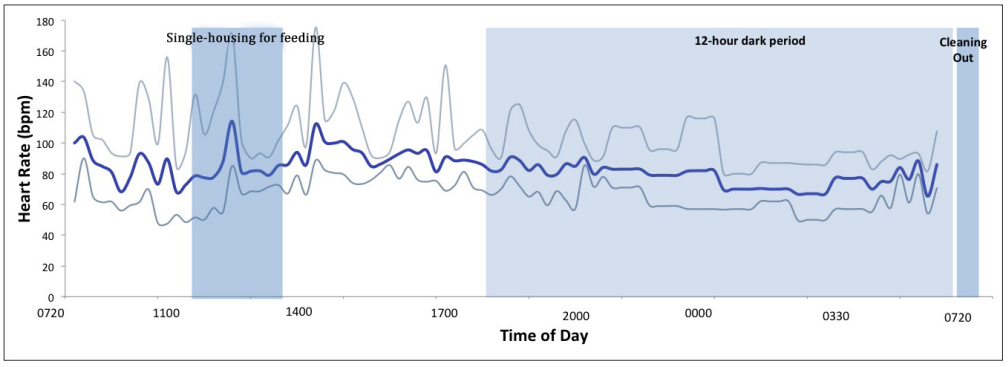


FIGURE C.3: Mean heart rate (bpm), with maximums and minimums over 24 hours

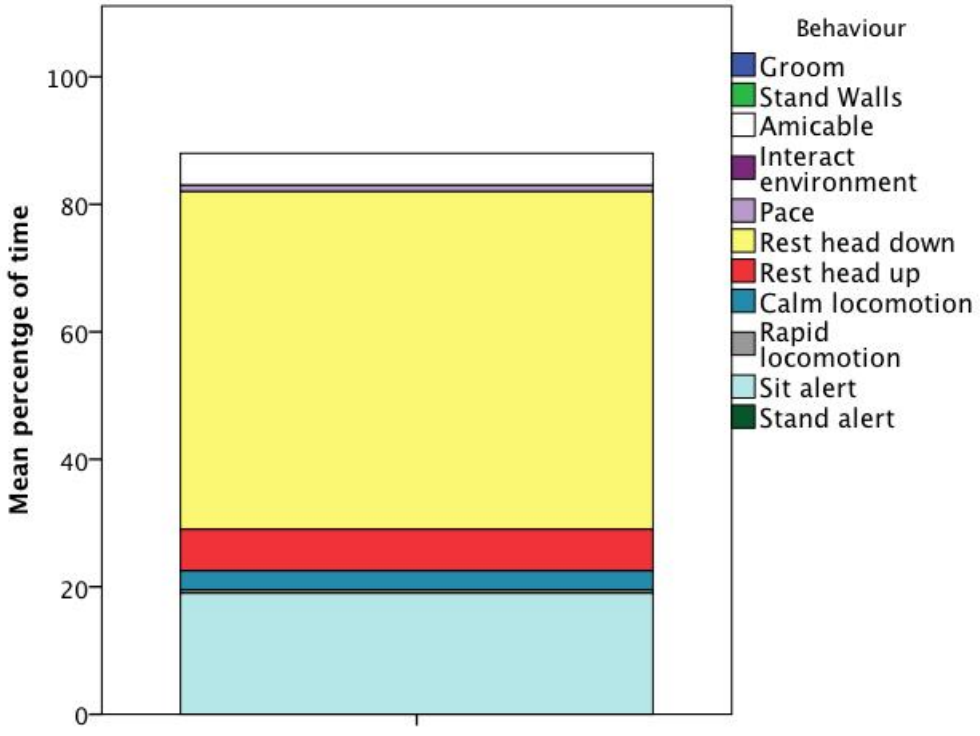


FIGURE C.4: Activity budget of six SP dogs over 24 hours

APPENDIX D

Data from Challenges, Chapter 6

D.1 Challenge 1: Human interaction

D.1.1 Changes in behaviour after human interaction

This appendix provides data to support the results of the study in Chapter 6. For each of the four challenges, the effects of time, group and affective state are shown, with analysis and graphical depictions.

TABLE D.1: Results of ANOVAs showing significant effects of time for HI

Behaviour	**F(3, 57)**	**p**
Interact with environment	7.013	<.001
Rest head down	7.493	<.001
Rest head up	31.661	<.001
Calm locomotion	5.000	.004
Sit alert	7.802	<.001
Stand alert	3.580	.019
High tail wag	126.144	<.001
Low tail wag	11.491	<.001
High posture	6.796	.001
Neutral posture	12.842	<.001
Half-low posture	4.297	.008

The effect of time for interacting with the environment is due to a decrease between After1 and After2; this is likely to reflect a decrease in exploration of scent cues following the end of the HI period. The effect of time of calm locomotion (p=.004) reflects

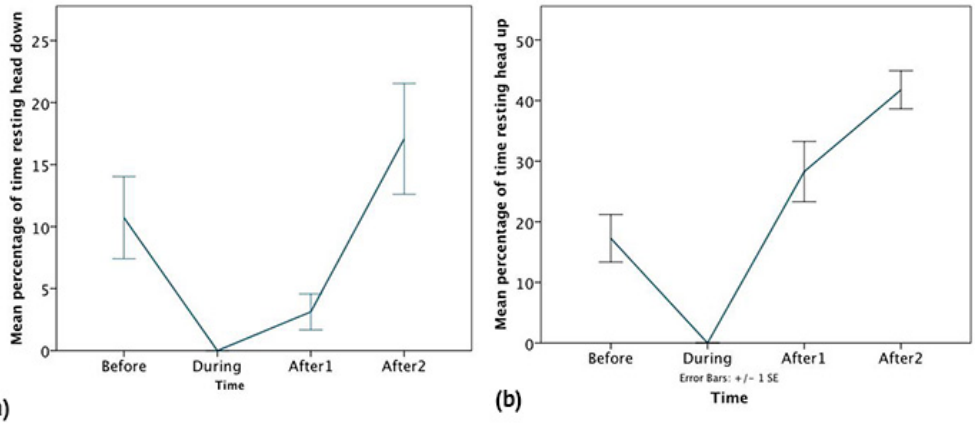


FIGURE D.1: Mean percentage of time (a) resting head down and (b) resting head up over time for HI

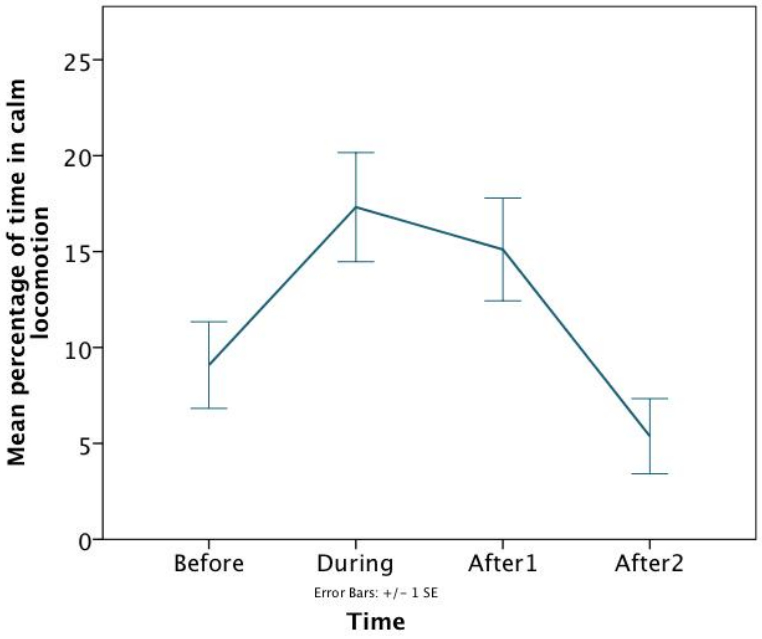


FIGURE D.2: Mean percentage of time in calm locomotion over time for HI

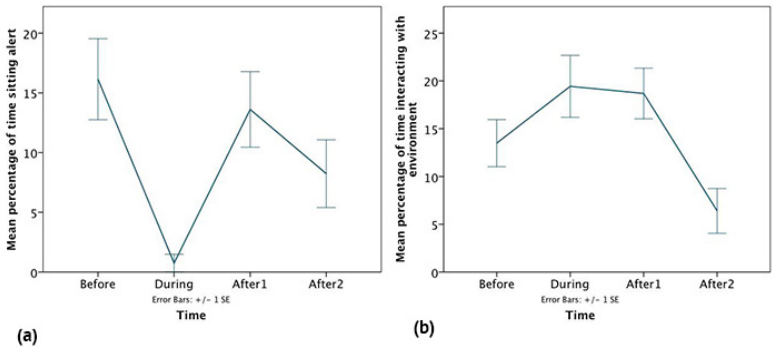


FIGURE D.3: Mean percentage of time (a) sitting alert and (b) standing alert over time for HI

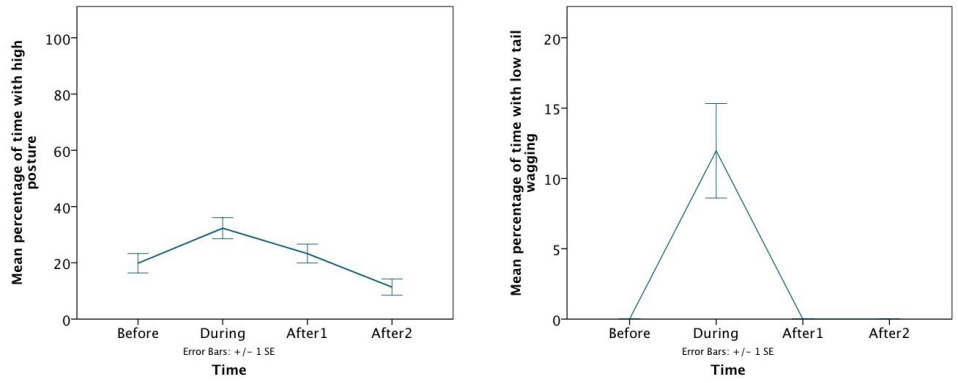


FIGURE D.4: Mean percentage of time with (a) high and (b) low tail wagging over time for HI

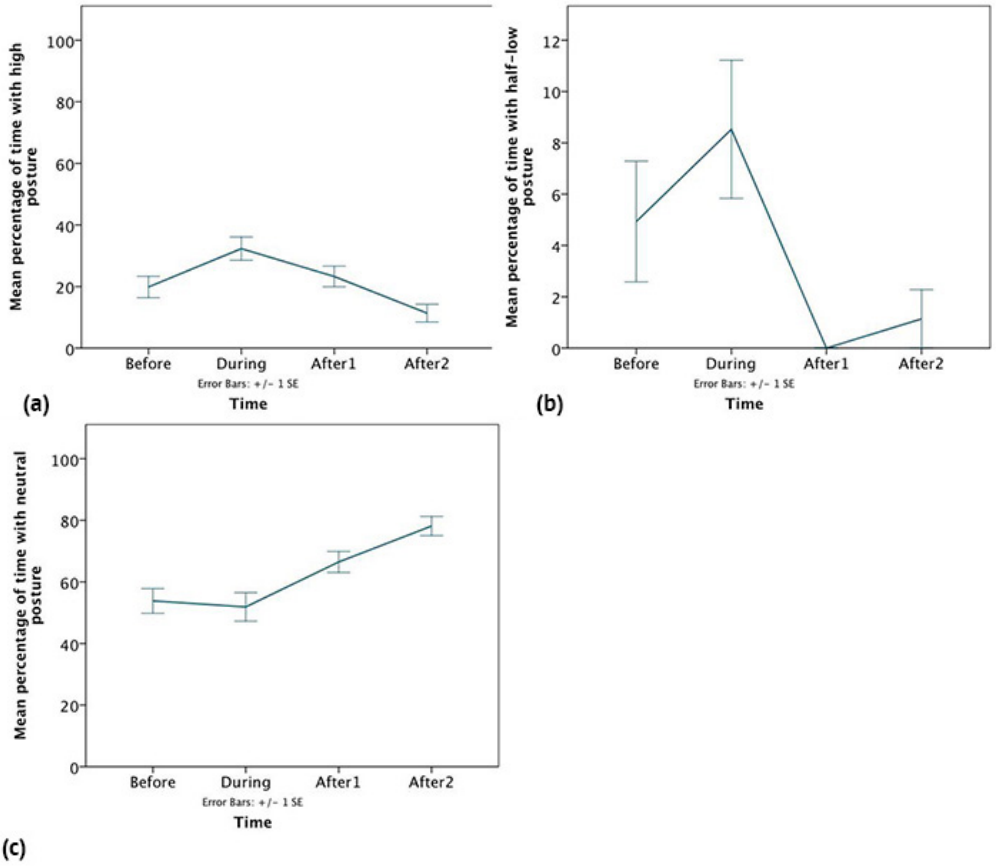


FIGURE D.5: Mean percentage of time in (a) high, (b) half-low and (c) neutral posture over time for HI

Table D.2: Post-hoc tests for changes in behaviour over time for HI

Behaviour	**Change**	**t(21)**	**p**
Interact with environment	After1 - After2	3.557	.02
Calm locomotion	After1 - After2	3.899	.001
Rest head down	Before - During	3.236	.004
	During - After1	5.690	<.001
	After1 - After2	3.822	.001
Rest head up	Before - After1	2.312	.031
	After1 - After2	3.268	.004
	Before - During	4.412	<.001
	During - After1	5.690	<.001
Sit alert	Before - During	4.607	<.001
	During - After1	3.786	.001
Stand alert	Before - During	2.344	.029
	During - After1	2.747	.012
High tail wag	Before - During	15.014	<.001
	During - After1	10.749	<.001
	After1 - After2	3.085	.006
Low tail wag	Before - During	3.565	.002
	During - After1	3.565	.002
High posture	Before - After 2	2.135	.045
	After1-After2	3.295	.003
Half-low posture	Before - After1	3.167	.005
	During to After1	2.097	.048
Neutral posture	Before - After1	3.077	.006
	During - After1 2.682	.014	
	After1 - After2	2.138	005

a decrease from After1 to After2. There was a NS increase from Before to During (p=.057). This is a similar pattern to interacting with environment and may also reflect an increase in exploration immediately after the HI ended.

The effect of time on resting head down (p<.001) was due to several changes in this behaviour between time points. There was no change from Before to After2. HI decreased resting head down but this effect did not last. There was also an effect of time on resting head up (p<.001). This is likely to be due to the above increase in activity During and at After1 and as with resting head down there was no lasting effect.

There were effects of time for sitting and standing alert (p<.001 and p=.019 respectively). Sitting alert decreased from Before to During increased again at After1; there was no difference from Before to After1 or from After1 to After2. The same was true for standing alert. While dogs did not exhibit these alert behaviours during HI (accounting for the decrease seen), this had no effect outwith the HI period.

There were effects of time for high and low tail wagging (p<.001 for both). As tail wagging only changed during the HI period, it is likely to be due to the presence of the experimenter. Low tail wagging is often used an an appeasement signal during social contact and its presence is likely to be related to attempts to interact with the experimenter. There were effects of time for high (p=.009), half-low (p=.003) and neutral posture (p=.008). HI appears to have immediate and longer-lasting effects on posture, with relaxed (neutral) posture increasing and undesirable postures (high and half-low) decreasing.

D.1.2 The effects of Group

Table D.3: Results of ANOVAs showing significant effects of Group for HI

Behaviour	**F(6, 57)**	**p**
Front of pen	3.011	.030
Interact with environment	8.015	<.001
High tail wag	2.880	.016
High posture	2.924	.005
Neutral posture	2.523	.031
Rest head up	3.227	.008
Half-low posture	3.166	.099

In addition to the effects on all dogs, human interaction had different effects on groups for time at front of pen, interacting with the environment, tail wagging and posture.

There was an interaction between time and Group for time spent at front of pen (p=.030). This is due to SP dogs showing an increase while DMPK and Stock dogs showed no change over time. Staff presence is greatly reduced in the SP dog area compared to the DMPK and Stock dog areas and is more frequently associated with regulated procedures so the increase in time at front immediately after HI may be due to increased arousal from excitement or anticipation.

Table D.4: Post-hoc tests showing the effects of Group on behaviour for HI

Behaviour	**Group**	**Change**	**t(7)**	**p**
Front	SP	Before - After1	4.265	.004
		After1 - After2	4.980	.041
Interact with environment	SP	During - After 1	5.277	.001
	Stock	Before - During	2.722	.042
High posture	SP	Before - During	6.506	<.001
		During - After1	2.580	.036
		After1 - After2	3.511	.011

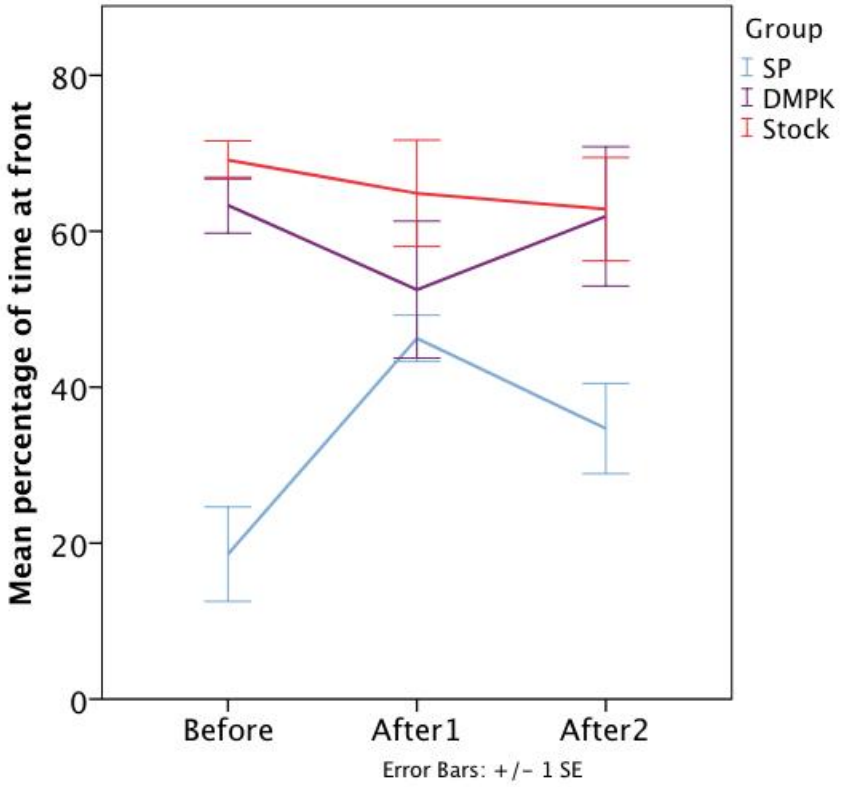


FIGURE D.6: Mean percentage of time at front of pen over time for HI

For interacting with the environment, the interaction between time and group ($p<.001$) was due to SP, but not DMPK or Stock showing an increase from During to After1 . Stock were the only group to show an increase from During the HI period. SP dogs were the only group to experience HI in the home pen so it is likely that the increase is due to scent cues while Stock dogs were the only group to experience HI in a play area also used by other groups of dogs so the increase is also likely to be due to scent cues.

The interaction between Group and time ($p=.016$) for high tail wagging is due to Stock dogs showing no difference in tail wagging from Before to After1 or After2 ($.111<p<.140$), i.e. tail wagging has returned to baseline levels immediately after the HI period, while it remained higher at After1 in the other groups. This may indicate a stronger response to HI in SP and DMPK. Stock dogs had a more stimulating environment.

The interaction between time and Group on high posture ($p=.005$) is due to SP dogs showing increases. Stock dogs showed no difference in high posture over time. This again suggests a stronger arousal response in SP and is related to increased time at the front. The interaction between time and Group on neutral posture ($p=.031$) is due to Stock showing no difference in neutral posture over the four time points. SP and DMPK showed the same increases as above. There were no effects of AS on behaviour before and after human interaction.

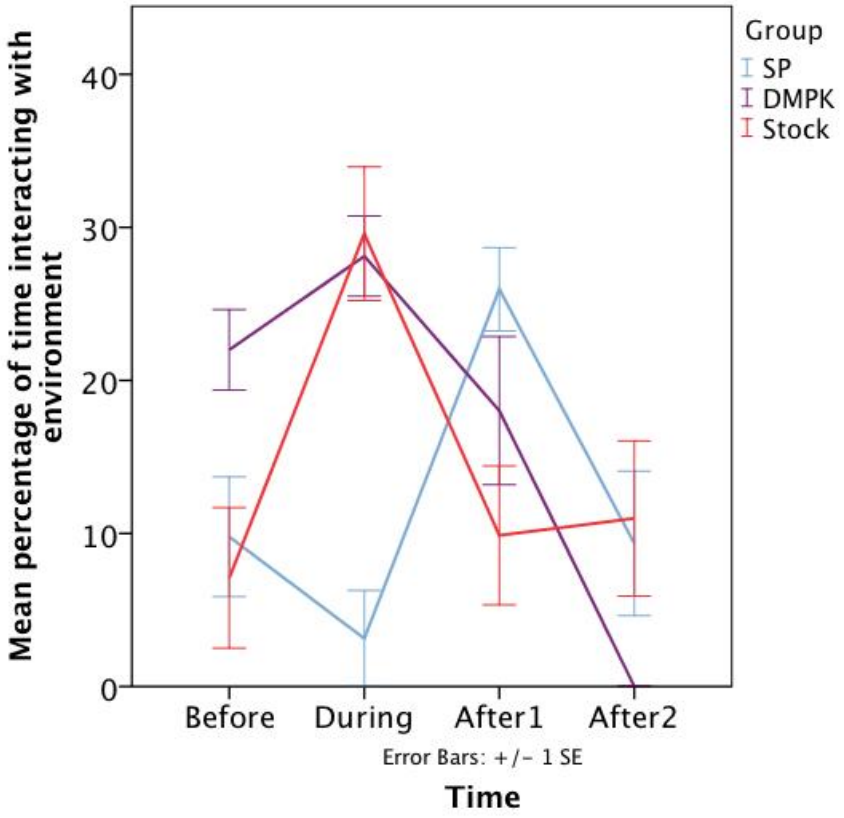


FIGURE D.7: Mean percentage of time interacting with environment over time and by Group for HI

D.2 Challenge 2: Single-housing

D.2.1 Changes in behaviour after single-housing

TABLE D.5: Results of ANOVAs showing significant changes in behaviour for single-housing

Behaviour	**F(3,32)**	**p**
Interact with environment	5.058	.001
Calm locomotion	4.088	.010
Sit alert	5.100	.003
Neutral posture	13.267	<.001
Half-low posture	10.876	<.001

Single-housing had effects on interacting in the environment, calm locomotion, alert behaviour and posture.

Interacting with the environment being was higher before and at After1 than During single-housing or at After2. Calm locomotion increased from Before to During; while

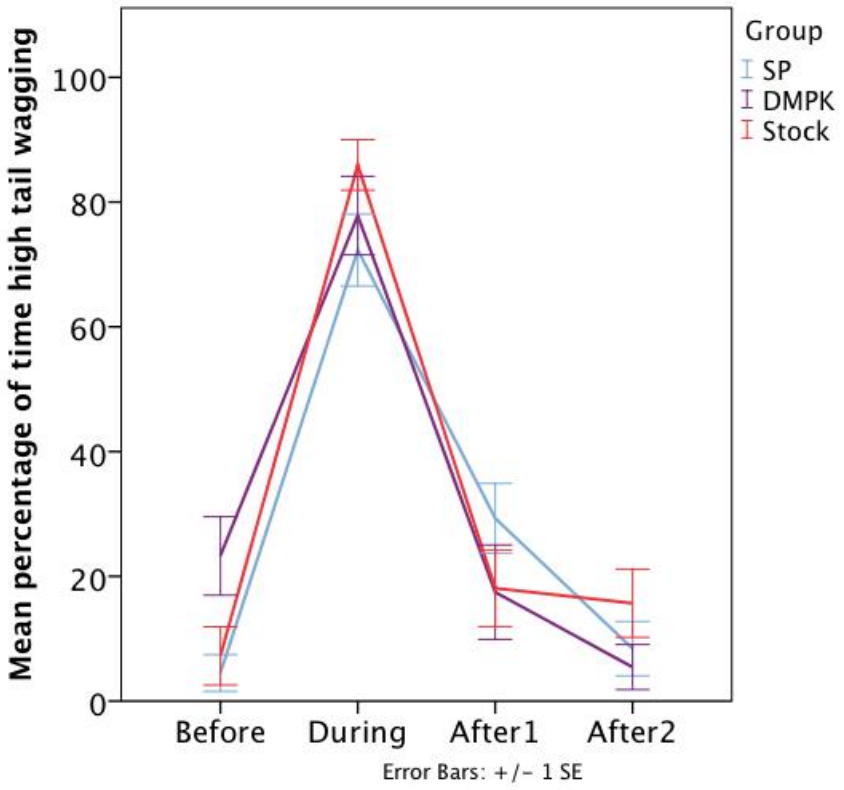


FIGURE D.8: Mean percentage of time high tail wagging over time and by Group for HI

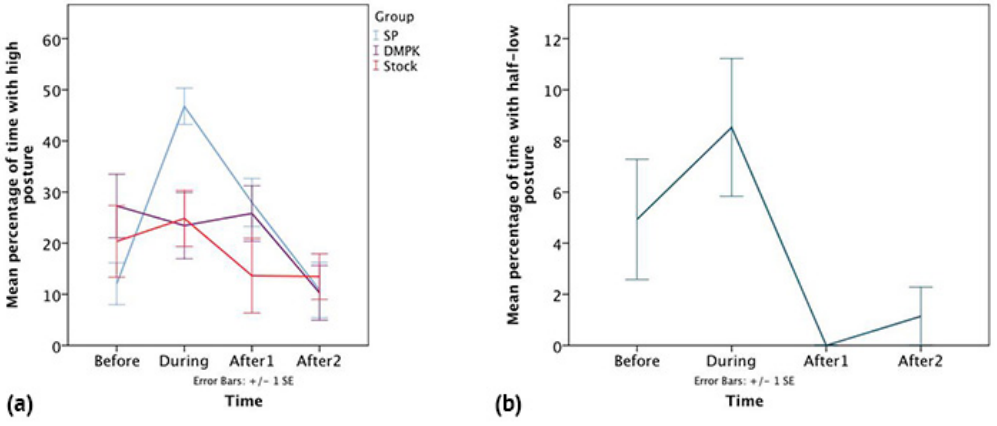


FIGURE D.9: Mean percentage of time with (a) high and (b) neutral posture over time and by Group for HI

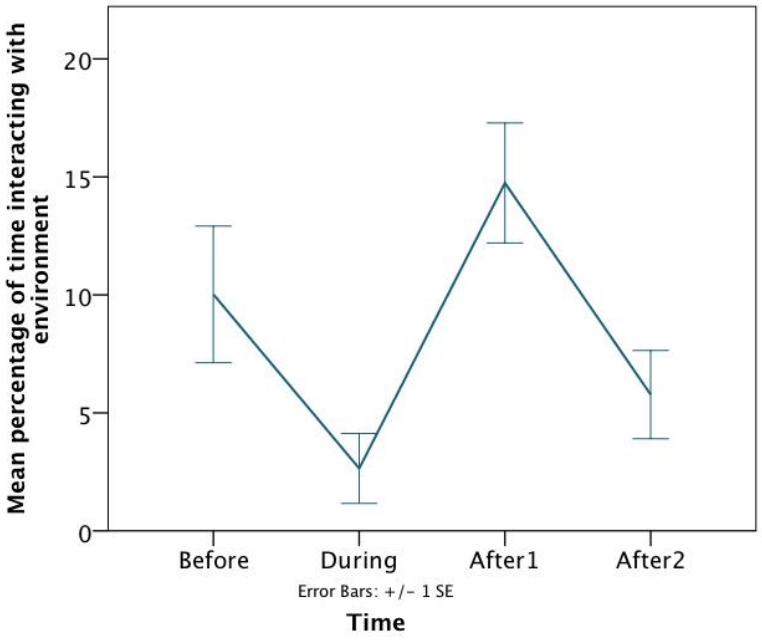


FIGURE D.10: Mean percentage of time interacting with environment over time for single-housing

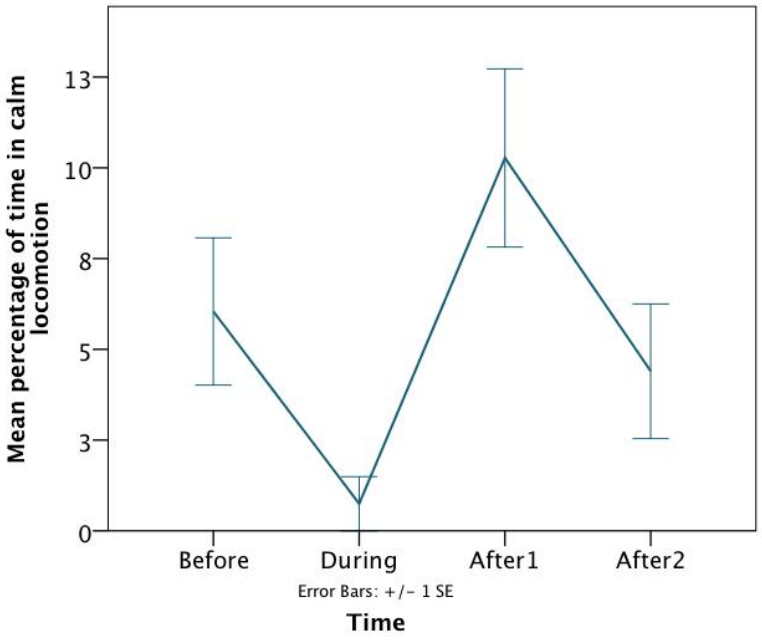


FIGURE D.11: Mean percentage of time spent in calm locomotion over time for single-housing

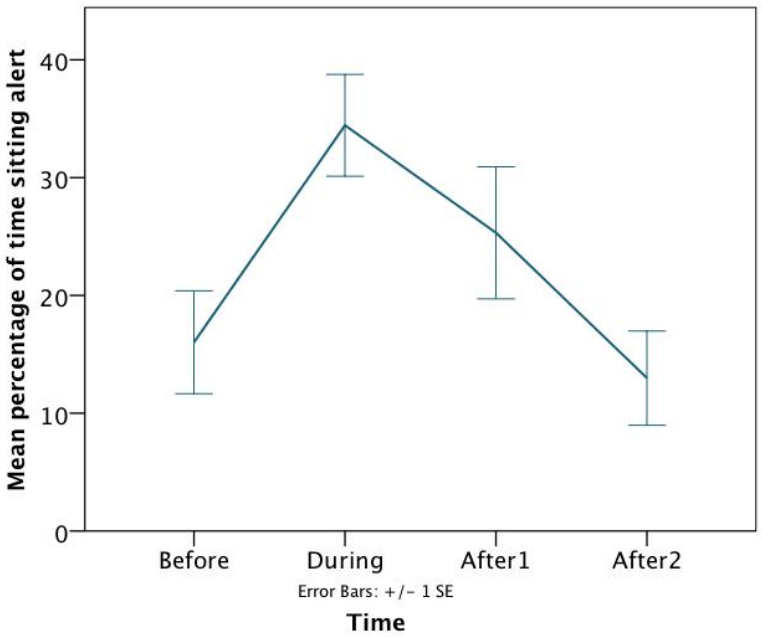


FIGURE D.12: Mean percentage of time spent sitting alert over time for single-housing

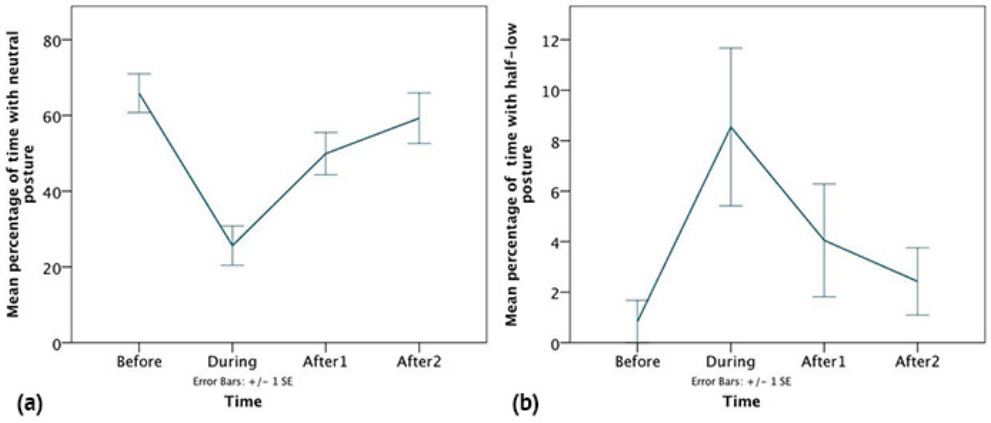


FIGURE D.13: Mean percentage of time spent with (a) neutral posture and (b) half-low posture over time for single-housing

sitting alert was also higher during single-housing. Neutral posture was lower after single-housing and half-low posture was higher during.

D.2.1.1 Group

There was no effect of Group on changes in behaviour in the home pen, suggesting that behaviour before and after a period of single-housing was a similar experience for all Groups.

Table D.6: Post-hoc tests for changes in behaviour over time for single-housing

Behaviour	Change	t(21)	p
Interact with environment	Before - During	2.840	.010
	During - After1	4.243	.007
	After1 - After2	2.963	.007
Calm locomotion	Before - During	2.345	.029
Sit alert	Before - During	4.123	<.001
	During - After2	3.686	.001
Neutral posture	Before - After1	2.694	.014
	During - After1	3.545	.002
	During - After2	4.210	<.001
Half-low posture	Before - During	2.666	.014
	During - After1	2.733	.012

D.2.1.2 Affective State

There was a marginally NS effect of AS on half-low posture at time point After1 (p=.056) with NAS dogs showing a trend towards more half-low posture.

The results of home pen behaviour analysis suggest that effect of single-housing is to decrease dogs' time interacting with the environment and with neutral posture, and increase time moving around the pen, sitting alert and with half-low posture. Neutral posture is still lower at After1 but has returned to previous levels by After2. These behaviours did not appear to be affected by Group or AS, suggesting that the effect of single-housing on behaviour in the home pen may be similar for all dogs.

D.3 Challenge 3: Feeding Toy

D.3.1 Changes in behaviour after Feeding Toy

Table D.7: Results of ANOVAs showing significant effects of time for FT

Behaviour	F(3, 48)	p
Interact with environment	9.977	<.001
Rest head don	7.954	<.001
Rest head up	10.830	<.001
Sit alert	13.700	<.001
High tail wag	69.432	<.001
High posture	3.242	.023
Neutral posture	6.389	.001
Events	4.376	.008

TABLE D.8: Results of Friedman's two-way analysis of variance by ranks showing significant effects of Group for FT

Behaviour	χ**2(2)**	**p**
Play	50.962	<.001
Panting	13.286	.004

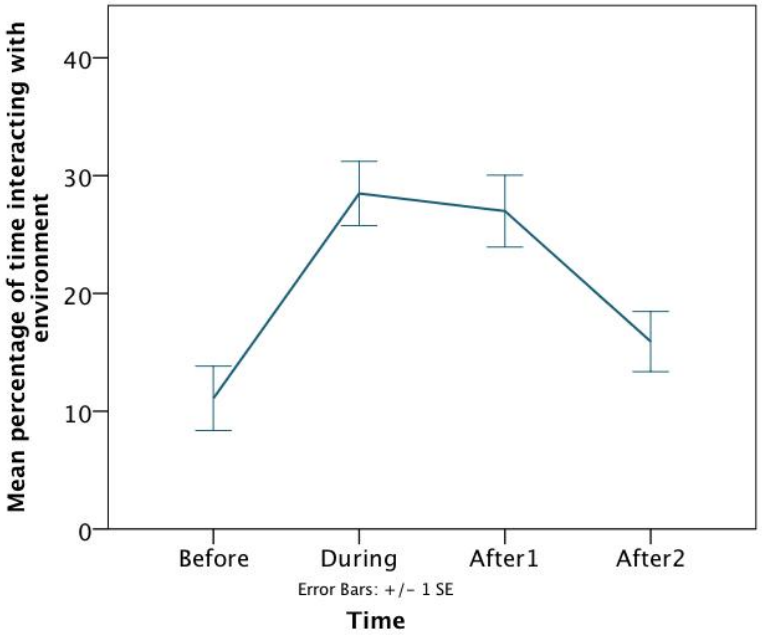


FIGURE D.14: Mean percentage of time interacting with environment over time for FT

Interacting with the environment is likely to have increased due to some food dropping onto the floor and remaining scent following the removal of the FT, although this appears to have no longer been present by After2.

There was an effect of time on resting head down (p<.001). Time resting head down decreased from Before to During (Z=2.366, p=.018). The increase in resting head up but not resting head down from Before to After2 suggests that dogs were not simply sleeping following the removal of the FT but were exhibiting more relaxed behaviour.

As the appearance of the FT was not a regular occurrence, increased sitting alert behaviour at After1 may be due to dogs attempting to observe staff, but this was not sufficient to increase it to above the level seen Before the FT was introduced and the FT has had the effect of reducing alert behaviour. The increase in tail wagging from Before to After1 may be due to dogs still interacting with the environment following the removal of the FT.

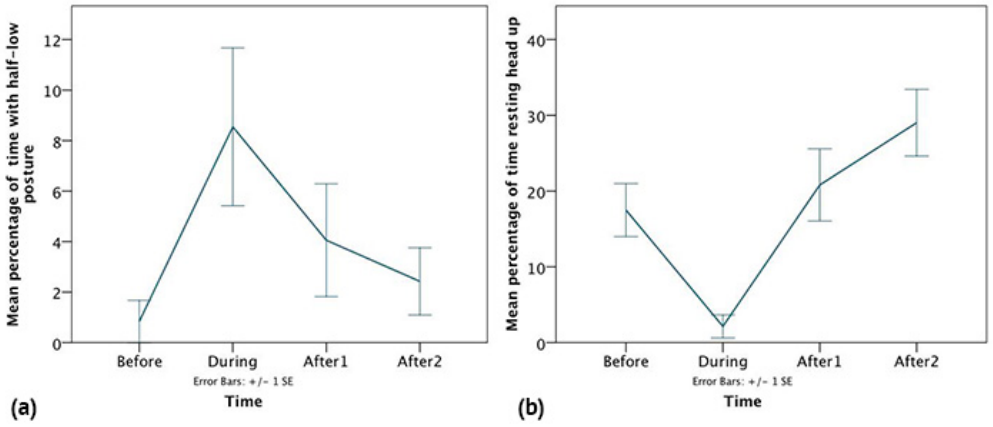


FIGURE D.15: Mean percentage of time (a) resting head up and (b) resting head down over time for FT

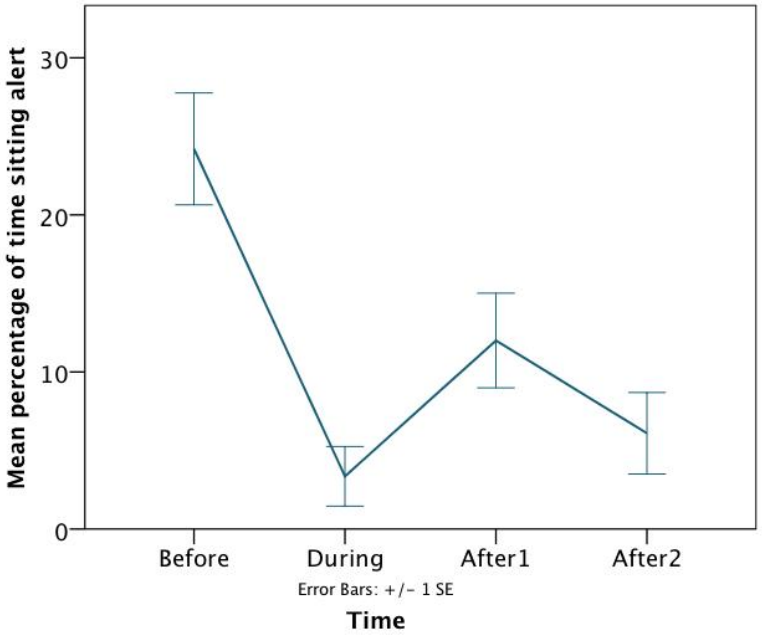


FIGURE D.16: Mean percentage of time sitting alert over time for FT

Taken together, the changes in high and neutral posture suggest that the FT did not cause high levels of excitement in the dogs but promoted an increase in relaxed posture. This effect was maintained for 30 minutes after the FT was removed.

The effect of time on total behavioural events (p=.008) is due to the rate of behavioural events decreasing from After1 to After2. There was a marginally NS increase from During to After1 (p=.051) and again from Before to After2 (p=.054). These behavioural events were predominantly oral behaviours and panting which are associated with food consumption as well as being stereotypic behaviours. The other predominant behavioural event was paw lifts which is a response to an unfamiliar or startling stimulus. As

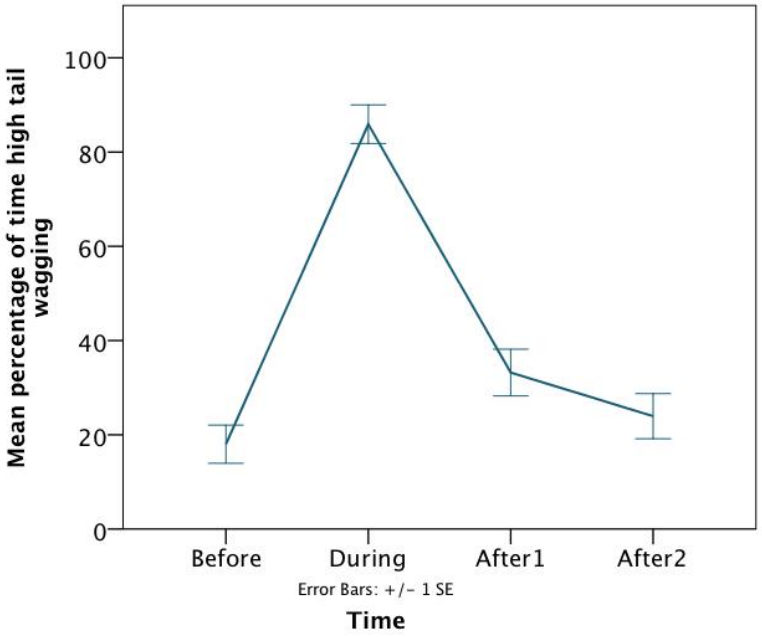


FIGURE D.17: Mean percentage of time high tail wagging over time for FT

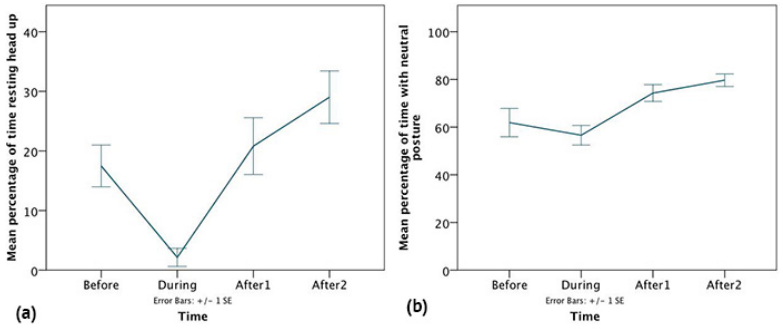


FIGURE D.18: Mean percentage of time with (a) high and (b) neutral posture over time for FT

the FT was not familiar to the dogs and was also interactive, this is an expected reaction to the FT.

There was an effect of time on play behaviour ($p<.001$) and panting ($p=.004$). Play increased from Before to During ($Z=4.002$, $p<.001$) but decreased again between During and After1 ($Z=4.020$, $p<.001$). Play was primarily directed at the FT during its presence in the pen but did not have an effect on other play behaviours following its removal. Panting increased from During to After1 ($Z=2.565$, $p=.010$), before decreasing again from After1 to After2 ($Z=2.266$, $p=.023$). However due to the increase in activity (play, interacting with environment) during the FT being present, the increase in panting immediately after it was removed is likely to be due activity and thirst rather than as a stereotypic behaviour.

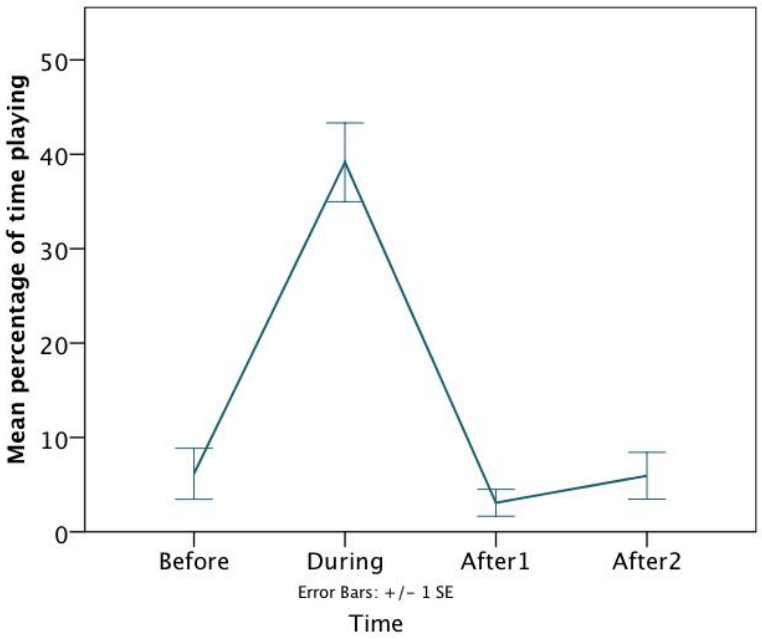


FIGURE D.19: Mean percentage of time playing over time for FT

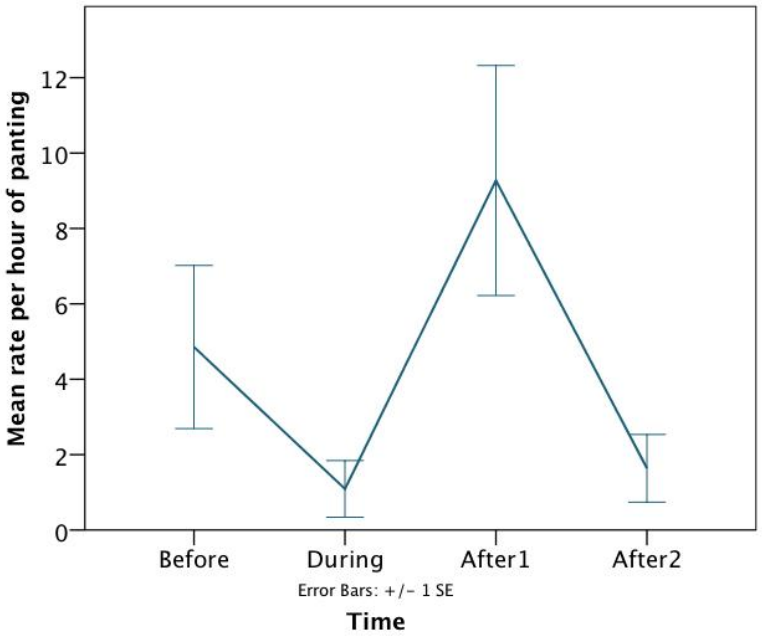


FIGURE D.20: Mean rate per hour of total events over time for FT

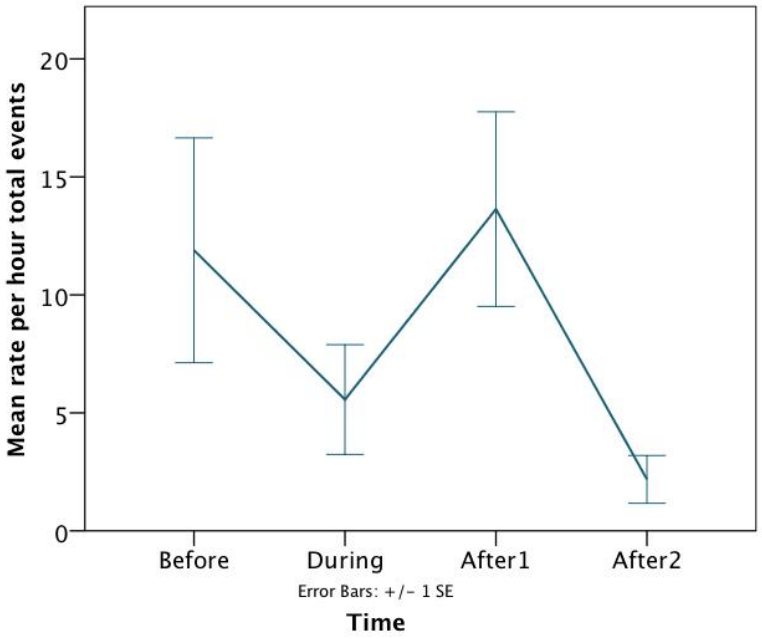


FIGURE D.21: Mean percentage of time interacting with environment over time for FT

TABLE D.9: Post-hoc tests showing changes in behaviour over time for FT

Behaviour	**Change**	**t(21)**	**p**
Interact with environment	Before - During	4.586	<.001
	During - After1	3.921	.001
	After1 - After2	3.063	.006
Rest head up	Before - During	3.738	.001
	During - After1	4.579	<.001
	Before - After2	3.395	.003
Sit alert	Before - During	5.600	<.001
	During - After1	2.579	.018
	Before - After1	3.632	.002
	Before - After2	3.395	.003
High tail wag	Before - During	11.234	<.001
	During - After1	8.202	<.001
	Before - After1	2.517	.020
Neutral posture	Before - After1	2.112	.047
	Before - After2	2.685	.015
High posture	Before - After2	2.637	.015
Events	After1 - After2	2.768	.012

D.3.2 The effects of Group

Table D.10: Results of ANOVAs showing significant effects of Group for FT

Behaviour	F(6, 48)	p
Sit alert	3.192	.009
Events	3.968	.002

Table D.11: Results of Kruskal-Wallis tests showing significant effects of Group for FT

Behaviour	Time point	H(2)	p
Barrier	After2	8.954	.011
Panting	Before	10.594	.005
Panting	After 1	7.048	.029

There was an effect of Group on time at the barrier at After2 due to SP dogs spending more time at the barrier than either DMPK (U=13.50, p=.042) or Stock (U=6.00, p=.010).

The interacting between time and Group for sitting alert (p=.019) is due to SP dogs showing no changes in sitting alert over any of the time points (.105<p<.992) while other dogs showed a decrease.

The interacting between Group and time for panting Before (p=.005) and After1 (p=.029) is due to no panting being observed for SP. This is likely to be an observation bias, as the position of the camera above the pen for SP (as opposed to in front of for DMPK and Stock) meant that it was difficult to observe the faces of the dogs.

The interaction between time and Group (p=.005) for total behavioural events is due to SP and Stock showing no change in the rate of behavioural events over time. In contrast, DMPK showed a decrease in behavioural events from Before to During , an increase from During to After2, a decrease from After1 to After2 and an overall decrease from Before to After2. This suggests that for DMPK, HI promoted more relaxed behaviour.

Table D.12: Post-hoc tests showing effects of Group on behaviour for FT

Behaviour	Group	Difference	t	df	p
Play	SP	>DMPK After2	2.643	14	.019
	SP	>Stock After2	2.266	12	.043
Events	DMPK	Before - During	2.865	7	.024
		During - After2	2.587	7	.036
		After1-After2	2.395	7	.022
		Before - After2	2.792	7	.027

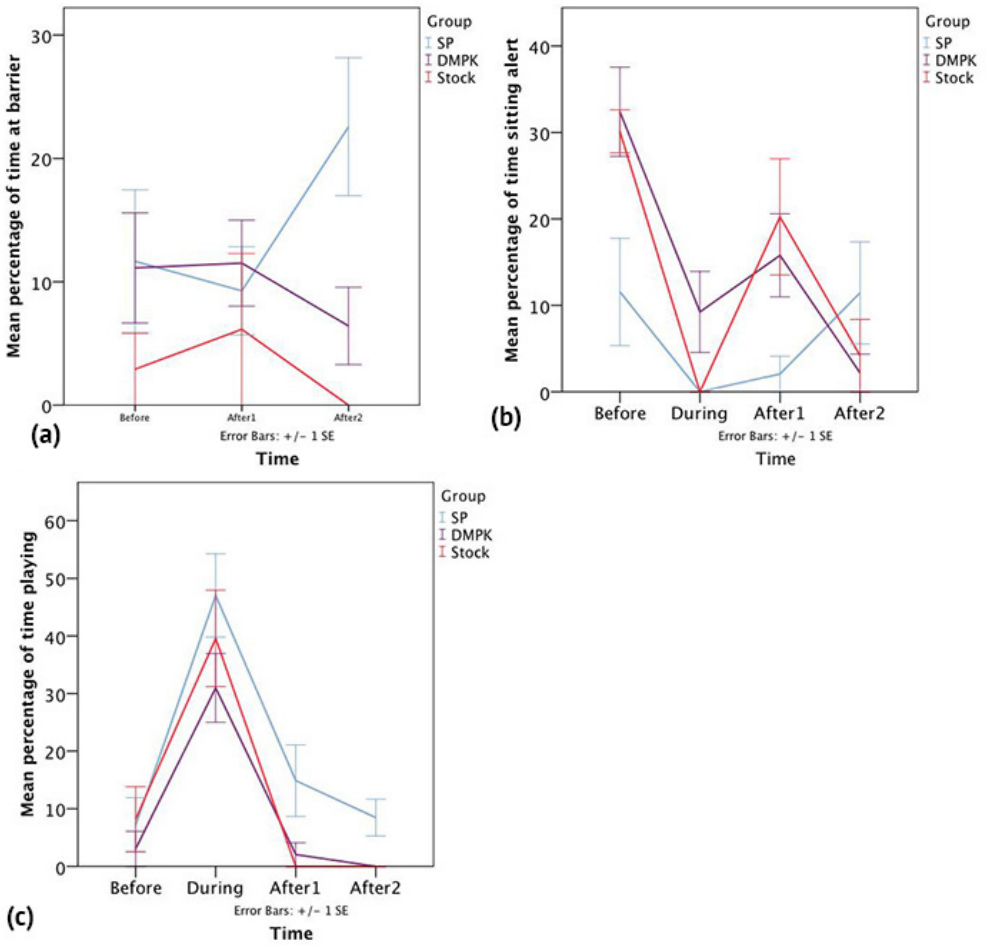


FIGURE D.22: Mean percentage of time (a) at barrier (b) sitting alert and (c) playing over time and by Group for FT

D.3.3 The effects of Affective State

The interaction between time and AS ($F(3, 48)=5.366$ $p=.003$) is due to NAS dogs spending more time interacting with the environment During ($t(20)=2.723$, $p=.013$) and at After2 ($t(20)=3.178$, $p=.005$) than PAS dogs. There were no other differences in behaviour between affective states.

D.4 Challenge 4: Restraint

D.4.1 Changes in behaviour after Restraint

Unlike other challenges, behaviour during Restraint could not be directly compared to home pen behaviour before or after as the nature of the challenge and change of location

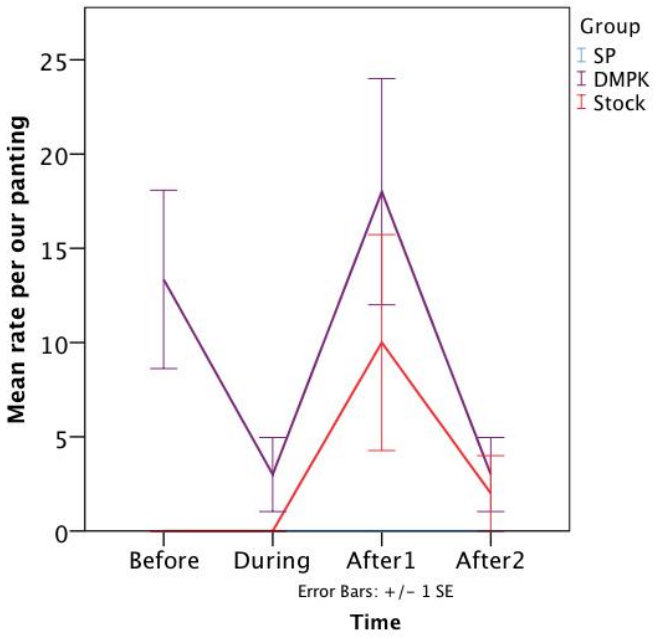


FIGURE D.23: Mean rate per hour panting over time and by Group for FT

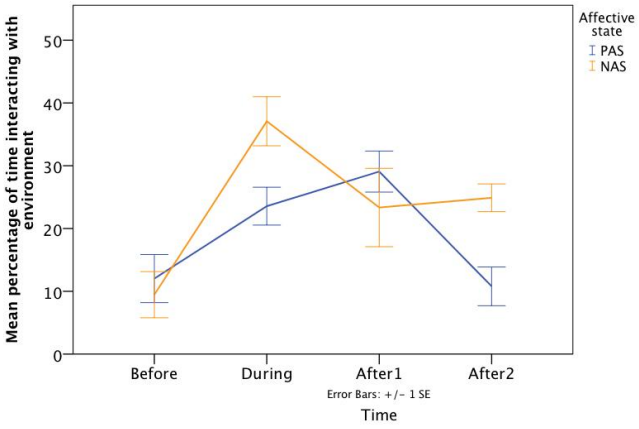


FIGURE D.24: Mean percentage of time interacting with environment over time and by Affective State

meant that many behaviours exhibited in the home pen could not be exhibited during restraint.

Resting head up decreased; as resting head up is a relaxed behaviour, restraint appears to have made the dogs less relaxed, while activity has increased. Behavioural events occurred at a high rate during restraint, but had returned to baseline levels immediately after. This suggests that behavioural events may be a more immediate response to a stressor rather than a long-lasting response.

There was a marginally NS effect of time on high tail wagging (p=.050), decreased

TABLE D.13: Results of ANOVAs showing significant effects of time for Restraint

Behaviour	**F(2, 32)**	**p**
Barrier	4.668	.017
Interact with environment	5.331	.010
Rest head up	3.825	.045
Events	5.1979	.011

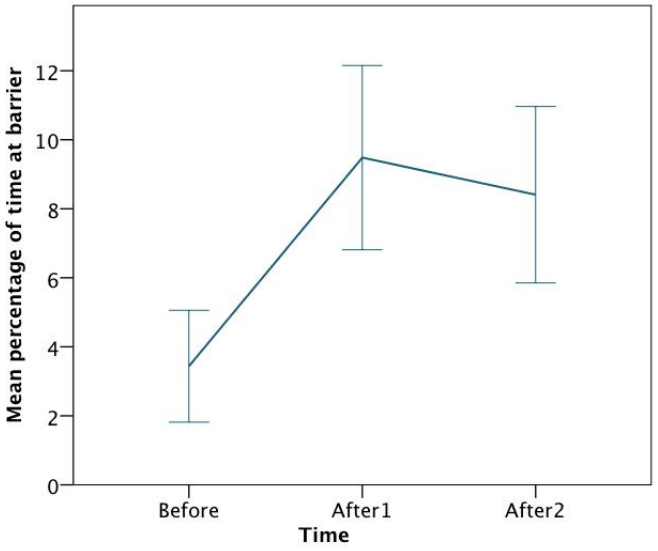


FIGURE D.25: Mean percentage of time at barrier over time for Restraint

TABLE D.14: Post-hoc analysis of changes in behaviour for Restraint

Behaviour	**Change**	**t(21)**	**p**
Barrier	Before - After1	2.662	.015
	Before - After2	2.345	.029
Interact environment	Before - After1	2.763	.012
	Before - After2	2.964	.007
Rest head up	Before -After1	2.459	.023
Events	Before - During	4.037	.001
	During - After1	4.962	<.001

during restraint and increased after. Tail wagging is an affiliative behaviour in a social context, a decrease during restraint suggests that dogs were not attempting to interact with the experimenter to ameliorate the negative experience, in contrast to the increase seen during HI.

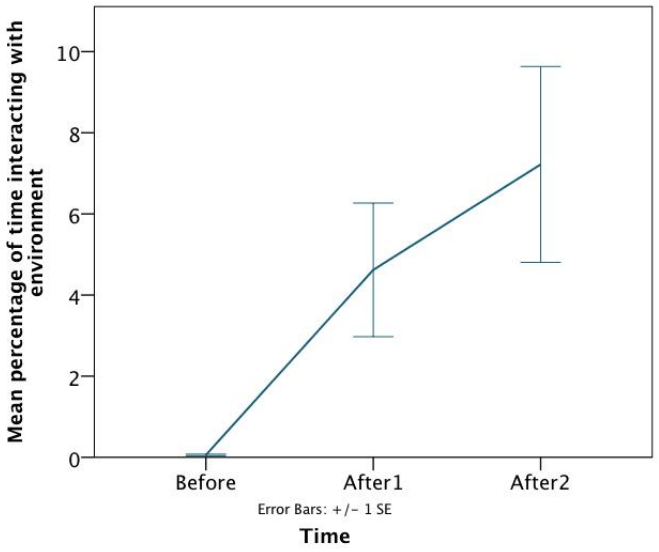


FIGURE D.26: Mean percentage of time interacting with environment over time for Restraint

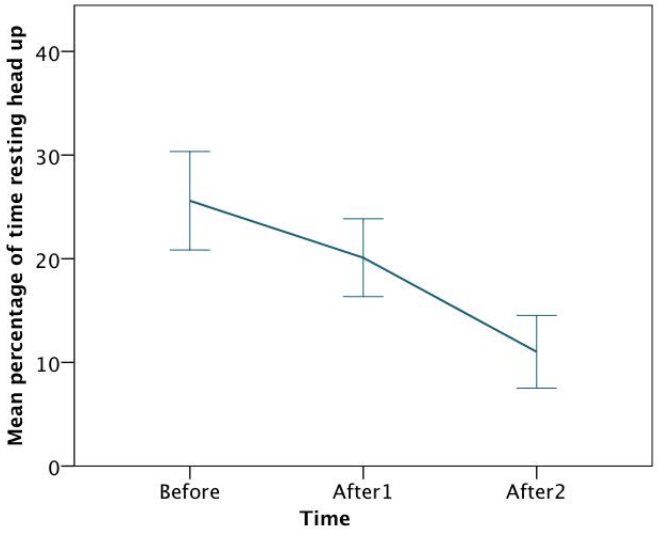


FIGURE D.27: Mean percentage of time resting head up over time for Restraint

D.4.2 The effects of Group on home pen behaviour before and after Restraint

Time at back of pen did not change for DMPK or Stock but increased for SP. SP showed an increase in time at barrier, while DMPK and Stock showed no changes. The decrease in time at the barrier is likely to be due to SP dogs spending more time at the rear of the pen at After2, following an increase due to agitation at After1. SP showed a decrease in amicable behaviour while DMPK and Stock showed no changes. Like resting behaviours, amicable behaviour reflects a relaxed state which suggests that

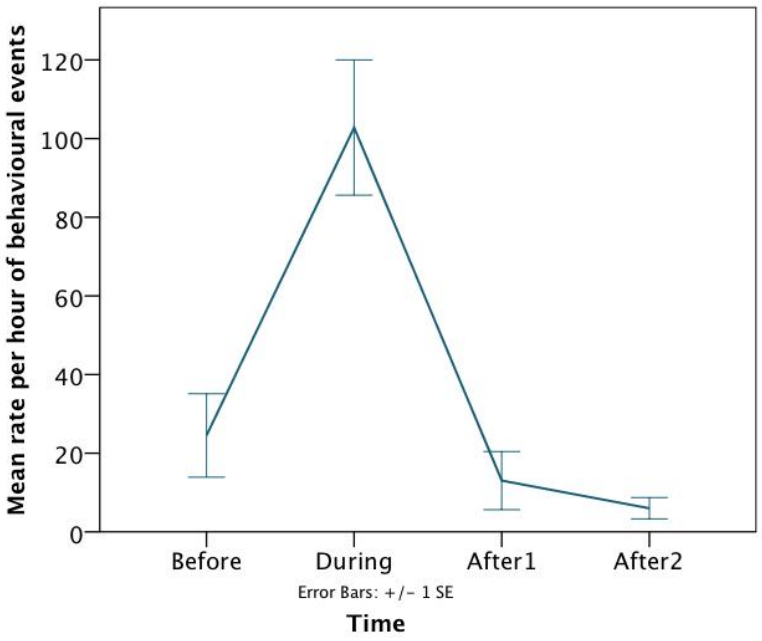


FIGURE D.28: Mean rate per hour of total events over time for Restraint

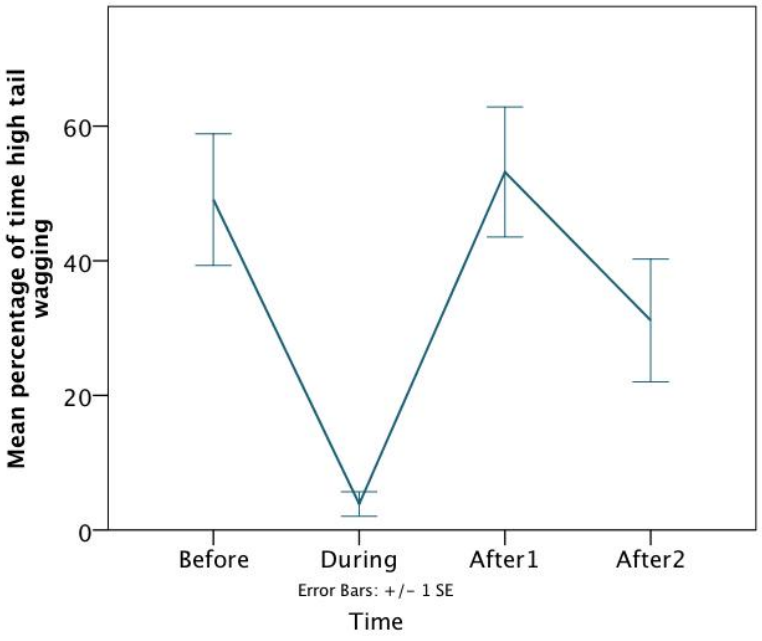


FIGURE D.29: Mean percentage of time high tail wagging over time for Restraint

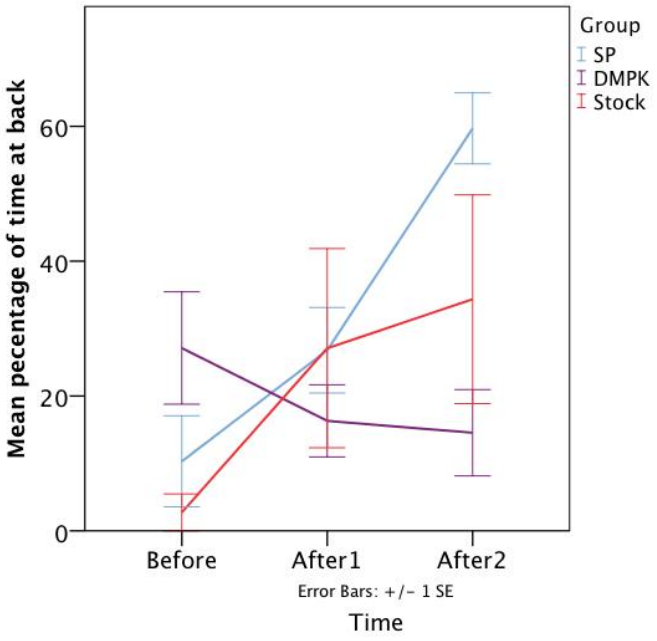


FIGURE D.30: Mean percentage of time at back of pen over time and by Group for Restraint

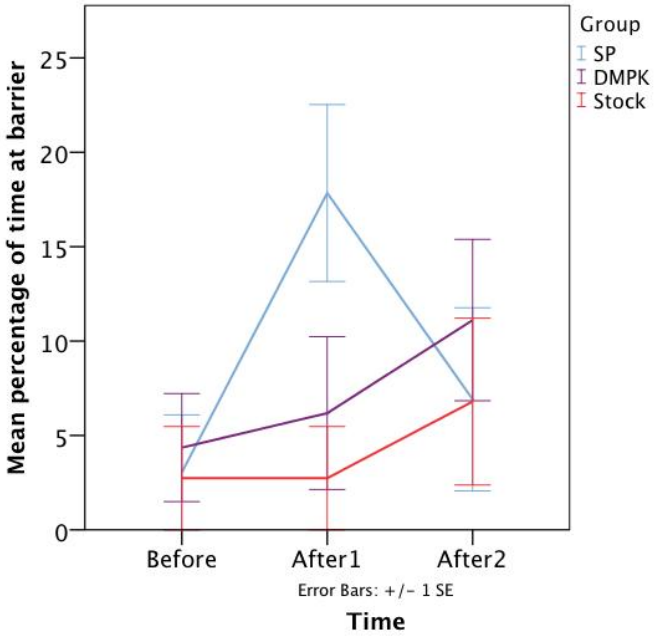


FIGURE D.31: Mean percentage of time at barrier over time and by Group for Restraint

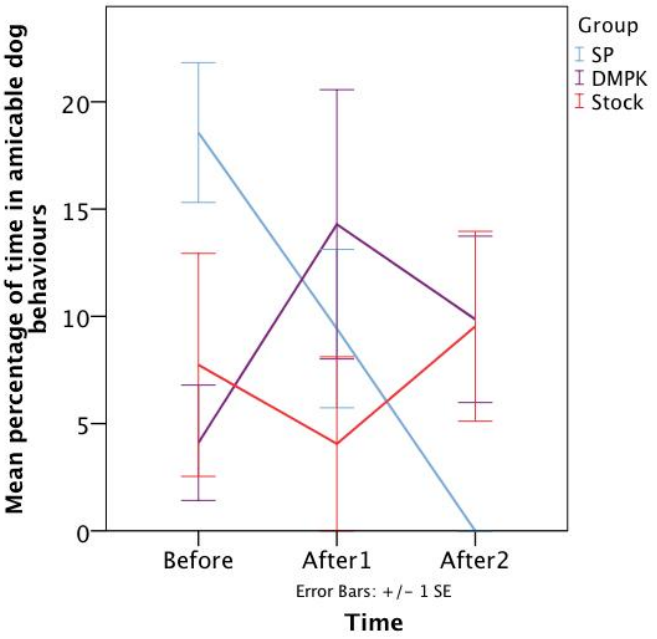


FIGURE D.32: Mean percentage of time in amicable behaviour over time and by Group for Restraint

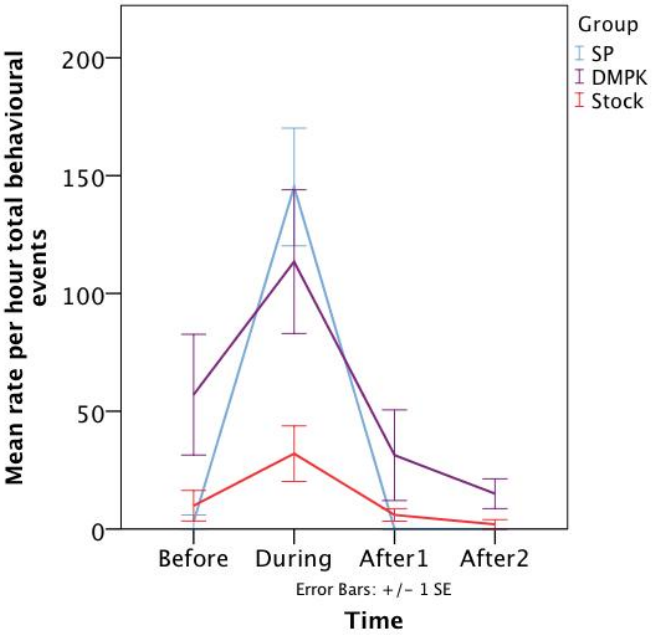


FIGURE D.33: Mean rate per hour total events over time and by Group for Restraint

Table D.15: Results of ANOVAs showing significant effects of Group for Restraint

Behaviour	F(4, 32)	p
Back of pen	3.075	.007
Barrier	4.041	.009
Amicable	5.062	.002
Events	3.113	.029

Table D.16: Post-hoc tests on the effects of Group on behaviour for Restraint

Behaviour	Group	Change	t(7)	p
Back of pen	SP	Before - After2	3.912	.006
		After1 - After2	4.421	.003
Barrier	SP	Before - After1	3.045	.019
Amicable	SP	Before - After2	5.697	.001
Events	Stock	Before - During	2.644	.046
	DMPK	During - After	2.455	.044
	SP	Before - During	5.831	.001
		During - After1	5.816	.001

restraint had the effect of reducing relaxed behaviours in SP dogs. Stock dogs showed a marginally significant increase in behavioural events from Before to During, while DMPK only showed a marginally significant decrease between During and After1. SP dogs showed a significant increase from Before to During and decrease from During to After1. This suggests that SP dogs had a stronger negative reaction to restraint than the other groups, although it was also aversive for DMPK and Stock.

D.4.3 Effects of Affective State on home pen behaviour before and after Restraint

Table D.17: Results of ANOVAs showing significant effects of Affective State for Restraint

Behaviour	F(2, 32)	p
Rest head down	3.688	.036
Stand alert	5.788	.007
Events	4.368	.021

PAS but not NAS dogs increased time resting head down from Before to After2. This suggests that PAS dogs were able to increase relaxed behaviours following restraint while NAS dogs were not. NAS dogs showed an increase in time standing alert between Before and After1 . Standing alert suggests that the NAS dogs may have been observing staff, as restraint frequently accompanies aversive events during company studies. NAS dogs

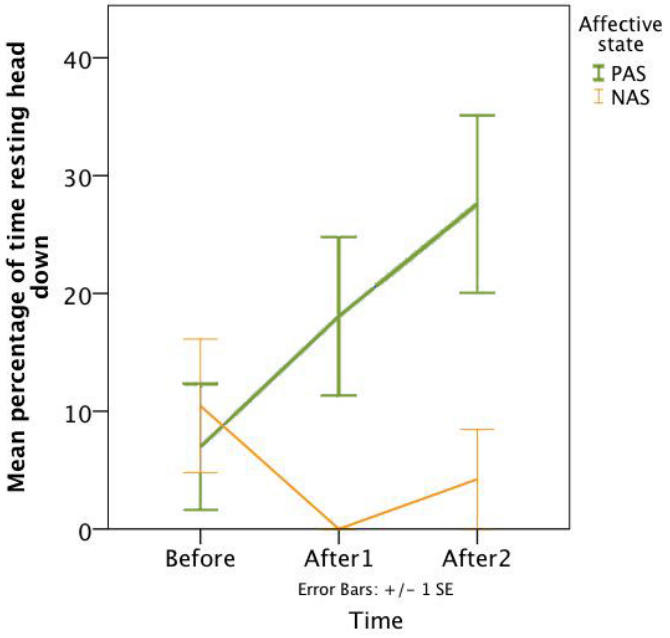


FIGURE D.34: Mean percentage of time resting head down over time and by Affective State for Restraint

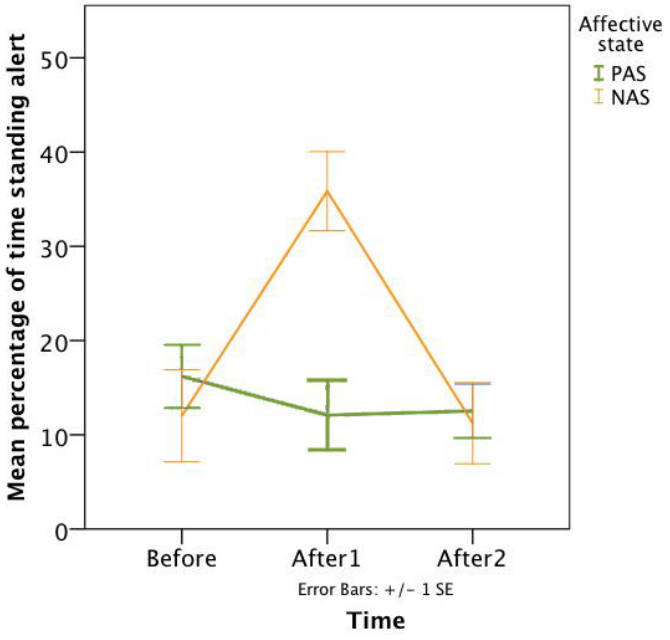


FIGURE D.35: Mean percentage of time standing alert over time and by Affective State for Restraint

TABLE D.18: Post-hoc tests for differences in behaviour by Affective State for Restraint

Behaviour	**AS**	**Change**	**t(7)**	**p**
Rest head down	PAS	Before - After2	2.351	.035
Stand alert	NAS	Before - After1	5.571	.001

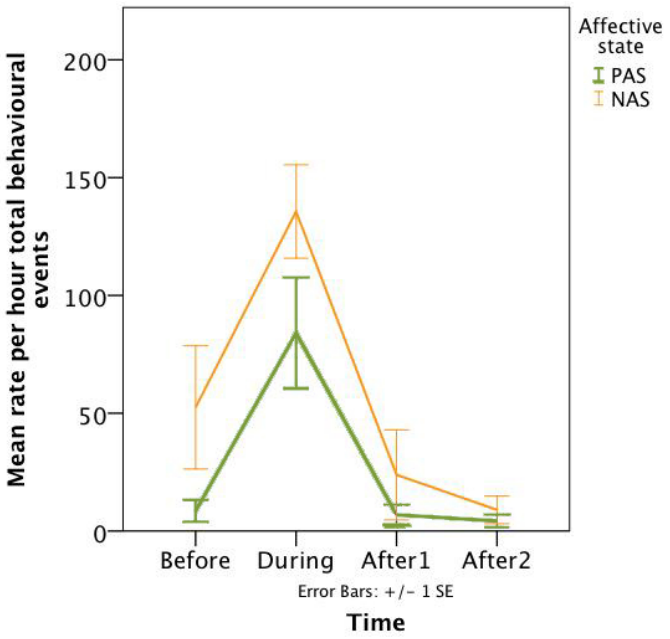


FIGURE D.36: Mean rate per hour total events over time and by Affective State for Restraint

also showed a NS increase in behavioural events from Before to During (p=.057) while PAS dogs did not. This suggests that NAS dogs may have had a stronger reaction to restraint than PAS dogs but this did not reach significance.

APPENDIX E

Technicians' Welfare Monitor Tool

The technicians' Welfare Monitoring Tool as used in Chapter 7 is presented below. The behaviours included were those which most were most strongly associated with positively- or negatively-valenced welfare in Chapters 5 and 6. A score of 0 was given to behaviours which had no clear welfare indication - calm moving and 'other' (to allow the technician to add in behaviours of interest which weren't included). Desirable behaviours were given a score of 1, while undesirable behaviours were given a score of 6 (moderately undesirable) or 10 (highly undesirable). Behavioural events and high or low posture were also given a score of 10. This score weighting meant that if a dog exhibited undesirable behaviours only once or twice throughout the day, scores would only be moderately increased, whereas dogs which exhibited undesirable scores throughout the day would have consistently high scores, in agreement with the Welfare Assessment Framework (Chapter 6). This differentiates between dogs exhibiting transitory changes in behaviour from those with consistent exhibition of negative welfare indicators.

Dog ID:		**Date:**					**Study Number:**			
Socially-housed (Y/N)										
Observed behaviours	**Behaviour**	**Score**	**8am**	**9am**	**10am**	**11am**	**12pm**	**1pm**	**2pm**	**3pm**
Activity level and behavioural states (score one only)	Other (add note)	0								
	Moving calmly	0								
	Resting – head up or head down	1								
	Interact with environment (explore, sniff)	1								
	Amicable dog interactions (sniff, groom, play)	1								
	Sitting alert	6								
	Stand against walls	10								
	Standing alert	6								
	Pacing between pens	10								
Posture (ear, body and tail position) (score one only)	Neutral (relaxed)	0								
	High (raised ears/body/tail)	10								
	Low (lowered ears/body/tail)	10								
Events (score all that apply)	Pant	10								
	Lip smacking or licking	10								
	Paw lift	10								
Total										
If dog is restrained for a procedure during that hour please also score below										
Additional behaviours to score during restraint (score all that apply)	Restrained without showing behaviour below	0								
	Very low posture	3								
	Crouch/tremble	4								
	Struggle/escape	4								
Total for restraint										
Initials										

Score range each hour: 0-75**.**

Interpretation: 0-25 no concern; 26-40 continue to monitor behaviour, consider intervention if score does not decrease; >40 welfare concern, intervention recommended if score consistently elevated.

Maximum score possible for restraint: 11**.** Score >1 intervention recommended.

APPENDIX F

Study plan for Chapter 7

The study plan for Chapter 7 is presented in this Appendix, providing details of the methodology. The principles of Good Statistical Practice (Peers et al., 2014) applied during this study are shown in Table F.1.

Table F.1: Good Statistical Practice Principle: Summary of Design

Appropriate type of design	A parallel group design was used
Appropriate control/ reference groups	The“control” group may be compared to the “sham” group to see the effect of sham dosing in weeks 1 and 2.
	The “sham” group may be compared to the “refined” group to see the effect of the additional features such as positive reinforcement, handling and predictability
Planned data analyses	Data to be analysed using appropriate factorial tests
Justification for animal numbers	A formal justification was not possible due to the lack of data on which to base power calculations. Six females per group seems like a reasonable minimum consistent with being able to detect meaningful group-to-group differences. The animals were returned to stock at the end of the study.
Blocking	Not applicable
Randomisation to treatment groups	Animals were assigned to groups to balance, as far as possible, the distribution of behaviours, ages and body weights Littermates will be dispersed through the groups
Appropriate processing order for treatment, sampling and termination Appropriate order for sample processing and analysis	One animal from group 1, followed by one animal from group 2, followed by one animal from group 3. Repeat 6 times. N/A
Blinding	The key behavioural endpoints are taken from observations of videos by LH who could not be blinded to the animal’s treatment group. This is a weakness, but no practical alternatives emerged.

Title	**Refining oral gavage in the beagle, and the comparison with sham dosing as a habituation technique**
Licensee	[redacted]
Study Director	[redacted]
Date of First Dose	27-May-13
Study End	31/05/2013
Project & Workbook Number	[redacted]
Procedure Number	2

Abbreviations used

BW	Body Weights	DO	Dosing
FD	Food	OB	Clinical Observations
PE	Physical examination	HR	Heart Rate

Insert additional instructions here:

There is no evening overtime

	Study Number : [redacted]								
	Compound : HPMC Vehicle								
	Unit 22B03			Unit 22B03			Unit 22B03		
	Control Group			**SD Group**			**RP Group**		
	DFD 27-May-13			DFD 27/05/2013			DFD 27/05/2013		
	No. animals **6F**			No. animals **6F**			No. animals **6F**		
Date	Wk No.	Day No.		Wk No.	Day No.		Wk No.	Day No.	
06-May-13	-1	-7	PE, BW	-1	-7	PE, BW	-1	-7	PE, BW
07-May-13		-6			-6			-6	
08-May-13		-5			-5			-5	
09-May-13		-4			-4			-4	
10-May-13		-3			-3			-3	
11-May-13		-2			-2			-2	
12-May-13		-1			-1			-1	

13-May-13	1	1	PE, BW, FD	1	1	PE, BW, FD	1	1	PE, BW, FD
14-May-13		2	OB, FD		2	OB, FD		2	OB, TRAIN, FD
15-May-13		3	FD		3	FD		3	OB, TRAIN, FD
16-May-13		4	OB, FD		4	OB, FD		4	OB, TRAIN, FD
17-May-13		5	FD		5	FD		5	FD
18-May-13		6	FD		6	FD		6	FD
19-May-13		7	FD		7	FD		7	FD
20-May-13	2	8	PE, BW, FD	2	8	PE, BW, FD	2	8	PE, BW, FD
21-May-13		9	OB, HR, MPTT, FD		9	HR, MPTT, FD		9	OB, TRAIN, HR, MPTT, FD
22-May-13		10	HR, MPTT, FD		10	OB, HR, MPTT, SHAM DO, FD		10	OB, HR, MPTT, SHAM DO, FD
23-May-13		11	OB, FD		11	OB, SHAM DO, FD		11	OB, SHAM DO, FD
24-May-13		12	FD		12	FD		12	FD
25-May-13		13	FD		13	FD		13	FD
26-May-13		14	FD		14	FD		14	FD
27-May-13	3	15	PE, BW, FD, DO, OB, HR, MPTT	3	15	PE, BW, FD, DO, OB, HR, MPTT	3	15	PE, BW, FD, DO, OB, HR, MPTT
28-May-13		16	DO, OB, FD, HR		16	DO, OB, FD, HR		16	DO, OB, FD, HR
29-May-13		17	DO, OB, FD, HR, MPTT		17	DO, OB, FD, HR, MPTT		17	DO, OB, FD, HR, MPTT
30-May-13		18	DO, OB, FD, HR		18	DO, OB, FD, HR		18	DO, OB, FD, HR
31-May-13		19	DO, OB, FD, HR, MPTT		19	DO, OB, FD, HR, MPTT		19	DO, OB, FD, HR, MPTT
01-Jun-13		20			20			20	
02-Jun-13		21			21			21	

APPENDIX G

Training plan for Chapter 7

Appendix 5 Training plan for Chapter 7

DAY 1 TITLE: First handling GOAL: Calm, relaxed removal from pen to procedure pod • Hand feeding • Calm approach • Holding and carrying • Removal to procedure pod • Interaction with handler must be +ve • Handler + procedure pod = +ve experience TROUBLESHOOTING: Dog nervous of handling/removal – remain in pen and work on building confidence with handler	DAY 2 TITLE: Table training GOAL: Calm on-table behaviour • Dog allowed to explore table + environment • Rewarded for calm behaviour • Equipment present – allow dog to investigate (incl. lubricant paste) TROUBLESHOOTING: dog nervous in PP – encourage exploration, reward for presence on table Dog excited – use verbal and hand signals to encourage dog to be stationary, reward
DAY 3 TITLE: First restraint GOAL: Dog is relaxed while gently restrained • Dog is comfortable on table and with handler • Gently restrain and reward for calm behaviour TROUBLESHOOTING: Dog nervous – reward for confident behaviour, reassure with vocal praise Dog excited – ignore excited behaviour, reward when calm, use vocal prompts if necessary	DAY 4 TITLE: Last restraint before SD GOAL: Dog is relaxed and neither nervous or excited when restrained AT PRESENT • Repeat 4 • Ensure dog is comfortable being relaxed, having mouth manipulated • Comfortable with presence of AT • Run through of SD protocol without performing SD TROUBLESHOOTING: nervous or excited dogs should have had sufficient training for relaxed behaviour, increased reward may be needed if not meeting criteria by this session
DAY 5 TITLE: First sham dose GOAL: gradual introduction of SD protocol, first SD +ve experience AT PRESENT • Gavage tube presented with lubricant paste, dog allowed to explore, licking/ chewing should be discouraged • Ensure dog calm and relaxed before SD • Dog should be praised and given extra reward for behaviour during SD TROUBLESHOOTING: note dogs which continue to be nervous, may need extra reward in Session 6.	DAY 6 TITLE: Second sham dose/final day of training GOAL: repeat of 5, ensure dog comfortable and relaxed before proceeding to dosing phase AT PRESENT • Repeat 5 • Dogs which were comfortable in 5 should again be given extra praise and reward for behaviour during SD • Dogs which remained nervous during 5 should be allowed extra time to become relaxed before SD TROUBLESHOOTING: note dogs which have not achieved desensitisation by this session for comparison during dosing

APPENDIX H

Implementation of the training protocol

Following analysis of the results of Chapter 7, the Industrial Partner decided to implement a protocol for all dogs currently held as stock (n=66, 27M + 39F). The training protocol was initially run over four weeks to determine the efficacy of the protocol as well as the ease of implementation for staff.

Over the period of four weeks, each dog received one training session using positive reinforcement training. The goal of the training was calm sitting while gently restrained. Trainers recorded both the stage of training achieved each week (A-G; Table ??) and behaviours exhibited (range -18 to +6; Table ??). All behaviours were scored as occurring not at all, some of the time or all of the time. At the end of the four weeks, the training stage and behaviour scores for each dog were analysed (Figure H.1). Seven dogs (3M + 4F) were identified as having little or no improvement in behavioural scores (range in Week 4 of -15 to -3) and training stages remaining at A-B. These dogs were given additional training (between one and six weeks), by the end of which all dogs had a minimum score of 0, however many remained at training stage A or B. From this, it appears that the response to training in the first four weeks predicted the success of training, those dogs which still had a negative score did not respond well to training.

Training scores for all dogs significantly improved from Week 1 to Week 4 (Wilcoxon signed ranks test $Z=6.483$, $p<.001$). Stages of training were at (range). Following an additional four weeks of training, the seven low-scoring dogs improved their behavioural scores to (range) and stage of training to (range).

Table H.1: Stages of training

Stage	Description
A	Doesn't accept treats (nervous or excited)
B	Accepts treats from handler
C	Calm and relaxed on table
D	Attempts sitting behaviour
E	Shows brief sits
F	Can maintain a longer sit
G	Sits well and tolerates gentle restraint

Table H.2: Behavioural indicators of welfare during training

Behaviour	Description	Score
Freeze	Rigid body, unreactive to the trainer	Some of the time -3 All of the time -6
Tremble or crouch	Physical shaking, body pressed into table	Some of the time - 3 All of the time -6
Escape attempts	Attempts to jump off table or remove itself from the trainer	Some of the time - 3 All of the time -6
Interact with trainer	Positive interactions with trainer e.g. eye contact, relaxed posture	Some of the time +3 All of the time +6

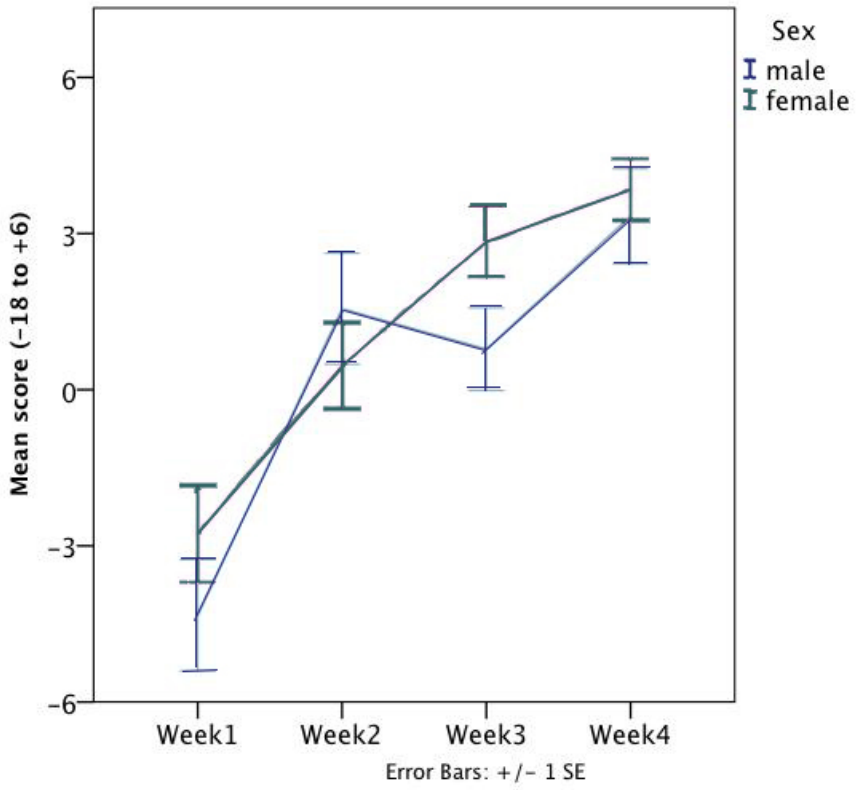


Figure H.1: Training scores over four weeks of training

APPENDIX I

Quality of science

Tables G.1-3 are taken from Tasker (2012). These tables provide a detailed description of the terms used to define aspects of what is termed "quality of science" in this thesis. Particular attention should be paid to "good measures" and "influential factors". Good measures are, in brief, measures which demonstrate the desirable qualities of the data obtained from animal use. While reproducibility and robustness were not tested, concerns about reliability (Chapter 5), sensitivity (Chapter 6) and repeatability (Chapter 6) are apparent.

Influential factors are those which can compromise good measures. Variation has been shown to be introduced by welfare state (Chapter 5) and by factors which were present between-groups such as husbandry, housing and regulated procedures. Welfare may also act as a confounding factor, causing differences in baseline blood pressure (Chapter 5), response to environmental stimuli (Chapter 6) and mechanical pressure threshold (Chapters 4 and 7).

TABLE I.1: Definition and explanation of terms used to describe quality of scientific outcomes part a

	Term	Definition		Explanation	Related & contributing factors	Protection, assumptions and methods of promotion
		Oxford Dictionary	Scientific			
Global terms	**Integrity**	The quality of being honest; lack of corruption.	The degree to which data collected and reported are what they purport to be (14).	**A combination of study management and scientific method quality.**	Validity Accuracy Precision Reliability Repeatability	Protected to a degree by compliance with GLP, following guidelines on study conduct and reporting by regulatory bodies (e.g. 17).
	Relevance	Closely connected to; Appropriate to the matter in hand.	The extent to which a test method correctly predicts or measures the biological effect of interest (12); appropriateness of tests and/or data for a particular hazard identification or risk characterisation (16).	The scientific basis to support the method (6); Establishing the scientific meaningfulness and usefulness of results for a particular purpose (1-4; 11 & 12).	Validity Accuracy Precision Reliability Repeatability	Prediction of toxicity is a complex process, dependent on: the selected animal model, the experimental design, data collection methods and the methods of extrapolation (13).
Right measures	Validity	*Valid* – actually supporting the intended point or claim.	The extent to which a measurement actually measures what the scientist wishes to measure and provides information to the questions being asked (16); Reliability and relevance of the method in supporting a specific use (2;3 & 17).	A valid or 'right' measure refers to the relationship between the variable under study and what it is supposed to measure or predict about the world. Valid measures are those that actually answer the questions being asked. Implies the experiment has a high probability of meeting the stated objectives (9); The objectives have a reasonable chance of contributing to human or animal welfare (9). Whether data derived can be generalised to other species- external validity (16)	Accuracy Specificity Relevance	Assumptions: (a) cynomolgus macaques are a relevant model; (b) the dose selection and route of administration match the intended target species (c) frequency and timing of data collection are based upon prior knowledge of drug action (d) core battery of biological variables are valid measures of organ function, histology, and pathology etc. Data must accurately represent the toxic endpoint being assessed (7).
	Accuracy	The degree to which the result of measurement, calculations, or specification conforms to the correct value or a standard.	Closeness to the real value (8); The degree to which measured/ calculated values reflect the true values of what they intend to represent (14); The closeness of agreement between a test result and an accepted reference value (12).	Is the measurement unbiased (free from systematic errors[a]), such that measured values correspond with true values? (16).	Validity Precision Specificity	Protected by compliance with quality assurance schemes (Section 2.4): instrumentation and analytical methods are accurate within, specified limits. Note reference values for comparison are influenced by environmental and animal factors.
	Specificity	Identify clearly and definitely; Precise and clearly repeatable.	To what extent does the measure describe what it is supposed to describe and nothing else i.e. the ability to detect true measures (16).	Is the measurement describing single or multiple: biological variable(s) and function(s); specific for pathology and free from interference from external variables?	Robustness Unwanted variation Confounding factors	Assume limitations of current analytical methods are known and specified.

TABLE I.2: Definition and explanation of terms used to describe quality of scientific outcomes part b

Good measures	**Reliability**	Consistently good in quality or performance.	The extent to which the measurement is repeatable and consistent (free from systematic errors[a]). Unbiased measurement, represents the true value of the variable, reduces the random component from imperfections in the measurement process (16). A reliable method produces results that are accurate and correctly reflect the sample being tested (15).	The smaller the error component the more reliable the measurement (16).	Precision Sensitivity Accuracy Precision	If the measurement is unreliable the real effects of the test article are difficult to quantify. This has implications for interpreting whether findings are biologically relevant i.e. are they bad? and risk posed prior to human exposure.
	Precision	The quality, condition or fact of being exact and accurate.	How free the measurements are from random errors[b] (16); Closeness of repeated measure to same value (8).	A measure of the reproducibility of the predictions of a model or repeated measurements (14).	Reproducibility Variation Accuracy	Reported for regulatory bodies.
	Sensitivity	Quick to detect or respond to slight changes, signals or influences.	The ability to reliably measure small changes, clearly distinguishable from background noise (14).	Are small changes in the true value reflected by changes in the measured value?(16)	Accuracy Precision Reliability	Depends on the sensitivity of the model to the toxicity of the test article – model and test article dependent. Lower and upper limits of determination are specified for analytical methods. Determining small changes in biological function may be impaired if the variation (background noise) is great or the physiological parameter is constrained by a ceiling or floor effects.
	Repeatability	Occur again in the same way and form.	The closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions with a given method (12); Can be described for measurements conducted repeatedly on a single individual or animals (within/intra-) or between individuals sampled at the same time in the same way (between/inter-).	The test result should be repeatable any number of times with low error (7).	Robustness Reproducibility Influential factors	Requirement for physiological parameters to be relatively stable over the given study time frame to enable a real difference in response to test article to be detected. Related to standardization, acclimatisation.
	Reproducibility	Produce a copy of, with a specified degree of success.	The degree to which a given method is reproducible within and between laboratories (12).	The biological endpoint must be reproducible. The same result is obtained if the test is repeated in other laboratories (7).	Repeatability Robustness	Most often used when developing alternative methods.

TABLE I.3: Definition and explanation of terms used to describe quality of scientific outcomes part c

	Robustness	Sturdy in construction; Strong, vigorous.	The insensibility of a test method to departures from the specified test conditions when conducted in different laboratories (12).	Robust method is one where successful results are obtained a high percentage of the time (13).	Specificity Reproducibility Repeatability	Most often used when developing alternative methods.
Influential factors	Variation	Change or slight difference in condition, amount, or level, typically within certain limits.	Term used to describe heterogenecity of values over time, space or different members of a given populations. An inherent property of the system or population (14).Intra-individual variability (difference over time in the same individual); inter-individual variability (differences between members of the population); Controllable variable –deliberately controlled for in the population, forms part of a control strategy in the method (14).	Any set of observations or measurements derived from a group of individuals will exhibit variability (1).	Accuracy Specificity Precision Repeatability Unwanted variation Confounding factor	Biological variation follows a normal distribution; source, handling, restraint and environmental conditions can affect variation in the population (e.g. confounding factor) – effects meaningful comparison to historic/ background data. Wanted variation describes the variation in response to test article. See below:
	Undesirable variation	*-undesirable variation* (variation as above)	Undesirable or unintended variation (16).	Skilful experimenter will attempt to eliminate these sources of variation (16).	Standardization	Use of a concurrent control group and comparable baseline data to identify undesirable variation.
	Confounding factor	*Confound-* Cause confusion; Mix up (something) with something else.	A confounding effect is caused by any other factor outside of the experimental treatment that might be present in the experimental arena (8).	Environmental or husbandry factors are examples of confounding factors that will impair ability to interpret study data	Standardization	Use of controls, randomization, and standard operating procedures to minimise the risk of confounding factors giving false negative or positive results. See standardization.
	Standardization	To conform to a standard; A required or agreed level of quality or attainment.	Setting standards; Defines properties of experimental animals and their environments to increase reproducibility of results (5); Keeping experimental conditions the same for all animals (within-experimental standardisation) or all experiments (between-experimental standardisation) (19).	Measures are taken to reduce variation relating to source, sex, environmental conditions, data capture and analysis by following standard operating procedures.		Trade off with external validity i.e. our ability to extrapolate results to other populations (Section 2.5). Previously used as an argument against group housing, environmental enrichment, socialisation and training programmes for laboratory housed animals.

[a]Systematic error: avoidable error due to controllable variables in a measurement; [b]Random error: Unavoidable errors that are always present in any measurement. Impossible to eliminate.
References (in alphabetical order): 1: Balls *et al* 1990a; 2: Balls *et al* 1990b; 3: Balls *et al* 1990c; 4: Balls *et al* 1995a; 5: Beynen *et al* 2001; 6: Bruner *et al* 1996; 7: Cronin 2005; 8: Dytham 2003; 9: Festing & Altman 2002; 10: Frazer 1990; 11: Gauch 2003; 12: ICCVAM 1997; 13: IPCS 1978; 14: IPCS 2008; 15: Klimisch *et al* 1997; 16: Martin & Bateson 2007; 17: OECD 1996; 18: Russell & Burch 1959; 19: Wurbel 2002.

References

Abbott, B. B., Schoen, L. S. & Badia, P. (1984). Predictable and unpredictable shock: behavioral measures of aversion and physiological measures of stress. *Psychological Bulletin*, *96*(1), 45.

Andics, A., Gacsi, M., Farago, T., Kis, A. & Miklosi, A. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology*, *24*(5), 574–578.

Animal Procedures Commitee. (2003). *Review of cost-benefit assessment in the use of animals in research.* London, UK: Home Office.

Animal Research Review Panel NSW. (2012). *Annual report 2011-2012.* Sydney, Australia: NSW Department of Primary Industries.

Badia, P., Culbertson, S. & Harsh, J. (1973). Choice of longer or stronger signalled shock over shorter or weaker unsignalled shock. *Journal of the Experimental Analysis of Behavior*, *19*(1), 25-32.

Badia, P., Harsh, J. & Abbott, B. (1979). Choosing between predictable and unpredictable shock conditions: Data and theory. *Psychological Bulletin*, *86*(5), 1107-1131.

Balcombe, J. P., Barnard, N. D. & Sandusky, C. (2004). Laboratory routines cause animal stress. *Journal of the American Association for Laboratory Animal Science*, *43*(6), 42–51.

Barnett, J. & Hemsworth, P. (1990). The validity of physiological and behavioural measures of animal welfare. *Applied Animal Behaviour Science*, *25*, 177-187.

Bassett, L. & Buchanan-Smith, H. M. (2007). Effects of predictability on the welfare of captive animals. *Applied Animal Behaviour Science*, *102*(3-4), 223-245.

Bateson, M., Desire, S., Gartside, S. E. & Wright, G. A. (2011). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology*, *21*(12), 1070-1073.

Batt, L., Batt, M., Baguley, J. & McGreevy, P. (2008a). The effects of structured sessions for juvenile training and socialization on guide dog success and puppy-raiser participation. *Journal of Veterinary Behavior: Clinical Applications and Research*, *3*(5), 199-206.

Batt, L., Batt, M. S., Baguley, J. A. & McGreevy, P. D. (2008b). Factors associated with success in guide dog training. *Journal of Veterinary Behavior: Clinical Applications and Research*, *3*(4), 143-151.

Bazett, H. C. (1920). An analysis of the time-relations of electrocardiograms. *Heart*, *7*, 353–370.

Beck, A. M. (1973). *The ecology of stray dogs: a study of free-ranging urban animals.* Purdue University Press.

Beden, S. N. & Brain, P. F. (1982). Studies on the effect of social stress on measures of disease resistance in laboratory mice. *Aggressive Behavior*, *8*, 126-129.

Beden, S. N. & Brain, P. F. (1984). Effects of attack-related stress on the primary immune responses to sheep red blood cells in castrated mice. *IRCS Medical Science: Psychology & Psychiatry*, *12*, 675.

Beden, S. N. & Brain, P. F. (1985). The primary immune responses to sheep red blood cells in mice of differing social rank or from individual housing. *IRCS Medical Science: Psychology & Psychiatry*, *13*, 364-365.

Beerda, B., Schilder, M., Bernadina, W., Van Hooff, J., De Vries, H. & Mol, J. (1999). Chronic stress in dogs subjected to social and spatial restriction. II. Hormonal and immunological responses. *Physiology and Behavior*, *66*(2), 243-254.

Beerda, B., Schilder, M., Van Hooff, J. & De Vries, H. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science*, *52*(3-4), 307-319.

Beerda, B., Schilder, M., Van Hooff, J., De Vries, H. & Mol, J. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, *58*(3-4), 365-381.

Beerda, B., Schilder, M., Van Hooff, J., De Vries, H. & Mol, J. (1999). Chronic stress in dogs subjected to social and spatial restriction. I. Behavioral responses. *Physiology and Behavior*, *66*(2), 233-242.

Beerda, B., Schilder, M. B., Janssen, N. S. & Mol, J. A. (1996). The use of saliva cortisol, urinary cortisol, and catecholamine measurements for a noninvasive assessment of stress responses in dogs. *Hormones and Behavior*, *30*(3), 272-279.

Benefiel, A. C., Dong, W. K. & Greenough, W. T. (2005). Mandatory enriched housing of laboratory animals: the need for evidence-based evaluation. *ILAR Journal*, *46*(2), 95-105.

Bentham, J. & Mill, J. S. (2004). *Utilitarianism and other essays.* Penguin UK.

Bergamasco, L., Osella, M. C., Savarino, P., Larosa, G., Ozella, L., Manassero, M., Badino, P., Odore, R., Barbero, R. & Re, G. (2010). Heart rate variability and saliva cortisol assessment in shelter dog: Human-animal interaction effects. *Applied Animal Behaviour Science*, *125*(1), 56–68.

Berns, G. S., Brooks, A. M. & Spivak, M. (2012). Functional MRI in awake unrestrained dogs. *PloS One*, *7*(5), e38027.

Bethell, E., Holmes, A., Maclarnon, A. & Semple, S. (2012). Cognitive bias in a non-human primate: husbandry procedures influence cognitive indicators of psychological well-being in captive rhesus macaques. *Animal Welfare*, *21*(2), 185.

Bhatnagar, S. & Meaney, M. J. (1995). Hypothalamic-pituitary-adrenal function in chronic intermittently cold-stressed neonatally handled and non handled rats. *Journal of Neuroendocrinology*, *7*(2), 97-108.

Blackwell, E.-J., Bodnariu, A., Tyson, J., Bradshaw, J. W. S. & Casey, R. A. (2010). Rapid shaping of behaviour associated with high urinary cortisol in domestic dogs. *Applied Animal Behaviour Science*, *124*(3), 113-120.

Blecha, F., Barry, R. A. & Kelley, K. W. (1982). Stress-induced alterations in delayed-type hypersensitivity to srbc and contact sensitivity to dnfb in mice. *Experimental Biology and Medicine*, *169*(2), 239-246.

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J. et al. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*, *92*(3), 375–397.

Boleij, H., Klooster, J., Lavrijsen, M., Kirchhoff, S., Arndt, S. & Ohl, F. (2012). A test to identify judgement bias in mice. *Behavioural Brain Research*, *233*(1), 45-54.

Bortolami, E., Murrell, J. C. & Slingsby, L. S. (2013). Methadone in combination with acepromazine as premedication prior to neutering in the cat. *Veterinary Anaesthesia and Analgesia*, *40*(2), 181-193.

Box, R. & Spielmann, H. (2005). Use of the dog as non-rodent test species in the safety testing schedule associated with the registration of crop and plant protection products (pesticides): Present status. *Archives of Toxicology*, *79*(11), 615-626.

Boxall, J., Heath, S., Bate, S. & Brautigam, J. (2004). Modern concepts of socialisation for dogs: implications for their behaviour, welfare and use in scientific procedures. *ATLA*, *32*, 81-93.

Bradshaw, J. (2011). *In defence of dogs*. London: Penguin UK.

Braithwaite, V. & Boulcott, P. (2007). Pain perception, aversion and fear in fish. *Diseases of Aquatic Organisms*, *75*(2), 131.

Branson, N. & Rogers, L. (2006). Relationship between paw preference strength and noise phobia in *Canis familiaris*. *Journal of Comparative Psychology*, *120*(3), 176.

Brayton, A. R. (1974). Effects of crowding on endocrine function and retention of the digenean parasite microphallus pygmaeus in male and female albino mice. *Journal of Helminthology*, *48*(02), 99-106.

Briefer, E. F. & McElligott, A. G. (2013). Rescued goats at a sanctuary display positive mood after former neglect. *Applied Animal Behaviour Science*, *146*(1), 45–55.

Brilot, B. O., Asher, L. & Bateson, M. (2010). Stereotyping starlings are more 'pessimistic'. *Animal Cognition*, *13*(5), 721-731.

Broom, D. (1986). Indicators of poor welfare. *British Veterinary Journal*, *142*(6), 524-526.

Broom, D. (1991a). Animal welfare: concepts and measurement. *Journal of Animal Science*, *69*(10), 4167-4175.

Broom, D. (1991b). Assessing welfare and suffering. *Behavioural Processes*, *25*(2), 117–123.

Broom, D. (2006). Behaviour and welfare in relation to pathology. *Applied Animal Behaviour Science*, *97*(1), 73-83.

Broom, D. & Johnson, K. (1993). *Stress and animal welfare.* Dordrecht: Kluwer Chapman and Hall.

Broom, D. & Kennedy, M. (1993). Stereotypies in horses: their relevance to welfare and causation. *Equine Veterinary Education*, *5*(3), 151-154.

Broom, D. & Kirkden, R. (2004). Welfare, stress, behaviour and pathophysiology. *Veterinary Pathophysiology*, 337-369.

Brown, M., Carbone, L., Conlee, K., Dawkins, M., Duncan, I., Fraser, D., Griffin, G., Hampshire, V., Lambert, L., Mench, J., Morton, D., Richmond, J., Rollin, B., Rowan, A., Stephens, M. & Wrbel, H. (2006). Report of the working group on animal distress in the laboratory. *Lab Animal*, *35*(8), 26-30.

Brydges, N. M., Leach, M., Nicol, K., Wright, R. & Bateson, M. (2011). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, *81*(1), 169-175.

Buchanan-Smith, H. M. (2006). Primates in laboratories: Standardisation, harmonisation, variation and science. *ALTEX: Alternatives to Animal Experiments*, *23*, 115-119.

Buchanan-Smith, H. M., Rennie, A., Vitale, A., Pollo, S., Prescott, M. & Morton, D. (2005). Harmonising the definition of refinement. *Animal Welfare*, *14*(4), 379-384.

Burman, O., McGowan, R., Mendl, M., Norling, Y., Paul, E., Rehn, T. & Keeling, L. (2011). Using judgement bias to measure positive affective state in dogs. *Applied Animal Behaviour Science*, *132*(3), 160–168.

Burman, O., Parker, R. M. A., Paul, E. S. & Mendl, M. (2008). Sensitivity to reward loss as an indicator of animal emotion and welfare. *Biology Letters*, *4*(4), 330-333.

Byrne, A. & Eysenck, M. W. (1993). Individual differences in positive and negative interpretive biases. *Personality and Individual Differences*, *14*(6), 849–851.

Canadian Council on Animal Care. (2012). *CCAC 2011 Animal Use Statistics.* Ottowa: CCAC.

Canli, T., Desmond, J. E., Zhao, Z., Glover, G. & Gabrieli, J. D. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *Neuroreport*, *9*(14), 3233–3239.

Chang, D. (2011). *Wigger's diagram.* Retrieved from http://en.wikipedia.org/wiki/File:Wiggers_Diagram.svg

Chapman, K., Sewell, F., Allais, L., Delongeas, J.-L., Donald, E., Festag, M., Kervyn, S., Ockert, D., Nogues, V., Palmer, H. et al. (2013). A global pharmaceutical company initiative: An evidence-based approach to define the upper limit of body weight loss in short term toxicity studies. *Regulatory Toxicology and Pharmacology*, *67*(1), 27-38.

Chen, H. (2007). The past to present animal use and current animal protection law in taiwan. *AATEX*, *14*, 743-748.

Christiansen, S. B. & Forkman, B. (2007). Assessment of animal welfare in a veterinary contexta call for ethologists. *Applied Animal Behaviour Science*, *106*(4), 203-220.

Chui, R. W., Fosdick, A., Conner, R., Jiang, J., Bruenner, B. A. & Vargas, H. M. (2009). Assessment of two external telemetry systems (physiojacket and jet) in beagle dogs with telemetry implants. *Journal of Pharmacological and Toxicological Methods*, *60*(1), 58–68.

Clough, G. (1988). Housing and welfare of laboratory rodents. In *New developments in biosciences: Their implications for laboratory animal science* (p. 239-244). Springer.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Coppola, C. L., Grandin, T. & Enns, R. M. (2006). Human interaction and cortisol: Can human contact reduce stress for shelter dogs? *Physiology & Behavior*, *87*(3), 537–541.

Corr, P. J., Wilson, G. D., Fotiadou, M., Kumari, V., Gray, N. S., Checkley, S. & Gray, J. A. (1995). Personality and affective modulation of the startle reflex. *Personality and Individual Differences*, *19*(4), 543–553.

Crabbe, J. C., Wahlsten, D. & Dudek, B. C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science*, *284*(5420), 1670–1672.

Damon, E., Eidson, A., Hobbs, C. & Hahn, F. (1986). Effect of acclimation to caging on nephrotoxic response of rats to uranium. *Laboratory Animal Science*, *36*(1), 24-27.

D'Andrea, I., Gracci, F., Alleva, E. & Branchi, I. (2010). Early social enrichment provided by communal nest increases resilience to depression-like behavior more in female than in male mice. *Behavioural Brain Research*, *215*(1), 71–76.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition*, *20*(1), 125–151.

Dawkins, M. (1990). From an animal's point of view: motivation, fitness, and animal welfare. *Behavioral and Brain Sciences*, *13*(1), 1-9.

Dawkins, M. (2004). Using behaviour to assess animal welfare. *Animal Welfare*, *13*(SUPPL.), S3-S7.

Dawkins, M. (2006). A user's guide to animal welfare science. *Trends in Ecology & Evolution*, *21*(2), 77–82.

Dawkins, M. (2008a). The science of animal suffering. *Ethology*, *114*(10), 937-945.

Dawkins, M. (2008b). The science of suffering. *Mental Health and Well-Being in Animals*, 47-55.

Dawkins, M. (2013). *The science of animal welfare.* London: Springer.

Dean, S. (1999). Environmental enrichment of laboratory animals used in regulatory toxicology studies. *Laboratory Animals*, *33*(4), 309-327.

Department of Primary Industries. (2012). *Statistics of animal use in research and teaching victoria, report number 29.* Victoria: Author.

Department of Primary Industries, Parks, Water and Environment. (2012). *Animal research statistics Tasmania, annual report number 16.* Department of Primary Industries.

Dettling, A. C., Feldon, J. & Pryce, C. R. (2002). Repeated parental deprivation in the infant common marmoset*(Callithrix jacchus)*, primates and analysis of its effects on early development. *Biological Psychiatry*, *52*(11), 1037-1046.

DeVries, A. C., Glasper, E. R. & Detillion, C. E. (2003). Social modulation of stress responses. *Physiology & Behavior*, *79*(3), 399–407.

Dickens, C., McGowan, L. & Dale, S. (2003). Impact of depression on experimental pain perception: a systematic review of the literature with meta-analysis. *Psychosomatic Medicine*, *65*(3), 369-375.

Dixon, M., Taylor, P., Slingsby, L., Hoffmann, M., Kästner, S. & Murrell, J. (2010). A small, silent, low friction, linear actuator for mechanical nociceptive testing in veterinary research. *Laboratory Animals*, *44*(3), 247-253.

Dixon, M., Taylor, P., Steagall, P., Brondani, J. & Luna, S. (2007). Development of a pressure nociceptive threshold testing device for evaluation of analgesics in cats. *Research in Veterinary Science*, *82*(1), 85-92.

Douglas, C., Bateson, M., Walsh, C., Bédué, A. & Edwards, S. A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Applied Animal Behaviour Science*, *139*(1), 65-73.

Doyle, R., Fisher, A., Hinch, G., Boissy, A. & Lee, C. (2010). Release from restraint generates a positive judgement bias in sheep. *Applied Animal Behaviour Science*, *122*(1), 28-34.

Doyle, R., Hinch, G., Fisher, A., Boissy, A., Henshall, J. & Lee, C. (2011). Administration of serotonin inhibitor p-chlorophenylalanine induces pessimistic-like judgement bias in sheep. *Psychoneuroendocrinology*, *36*(2), 279-288.

Doyle, R., Lee, C., Deiss, V., Fisher, A., Hinch, G. & Boissy, A. (2011). Measuring judgement bias and emotional reactivity in sheep following long-term exposure to unpredictable and aversive events. *Physiology and Behavior*, *102*(5), 503-510.

Dwyer, C. (2004). How has the risk of predation shaped the behavioural responses of sheep to fear and distress? *Animal Welfare*, 269-285.

European Economic Community. (1986). The approximation of laws, regulations and administrative provisions of member states regarding the protection of animals used for experimental and other scientific purposes. *Official Journal of the European Union*, *L358*, 1-28.

European Union. (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union*, *L 276/33*.

Europen Commission. (2013). *Seventh report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the european union.* Brussels: European Commission.

Even, N., Devaud, J.-M. & Barron, A. B. (2012). General stress responses in the honey bee. *Insects*, *3*(4), 1271–1298.

Everaert, J., Koster, E. H. & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review*, *32*(5), 413-424.

Everds, N. E., Snyder, P. W., Bailey, K. L., Bolon, B., Creasy, D. M., Foley, G. L., Rosol, T. J. & Sellers, T. (2013). Interpreting stress responses during routine toxicity studies a review of the biology, impact, and assessment. *Toxicologic Pathology*, *41*(4), 560–614.

Eysenck, H. J. (1991). *Dimensions of personality.* New York, U.S.A: Springer.

Fallani, G., Prato Previde, E. & Valsecchi, P. (2007). Behavioral and physiological responses of guide dogs to a situation of emotional distress. *Physiology & Behavior*, *90*(4), 648–655.

Festing, M. (2010). The design of animal experiments. In R. Hubrecht & J. Kirkwell (Eds.), *The UFAW Handbook on the Care and Management of Laboratory and Other Research Animals: Eighth Edition* (p. 23-36).

Festing, M. & Altman, D. (2002). Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR journal*, *43*(4), 244–258.

Fjellanger, R., Andersen, E. & McLean, I. (2002). A training program for filter-search mine detection dogs. *International Journal of Comparative Psychology*, *15*(4), 278-287.

Flannigan, G. & Dodman, N. H. (2001). Risk factors and behaviors associated with separation anxiety in dogs. *Journal of the American Veterinary Medical Association*, *219*(4), 460–466.

Fogel, B. (1990). *The dog's mind: Understanding your dog's behavior.* New York: Macmillan.

Fox, M. & Stelzner, D. (1966). Behavioural effects of differential early experience in the dog. *Animal Behaviour*, *14*(2), 273-281.

Fraser, A. F. & Broom, D. M. (1990). *Farm animal behaviour and welfare.* Oxon: CaB international.

Fraser, D. (2008). Understanding animal welfare. *Acta Veterinaria Scandinavica*, *50*(SUPPL. 1).

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alka, C., Vil, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., Lee, C., Tadigotla, V., Siepel, A., Bustamante, C. D., Harkins, T. T., Nelson, S. F., Ostrander, E. A., Marques-Bonet, T., Wayne, R. K. & Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics*, *10*(1), e1004016.

Gacsi, M., Gyori, B., Miklosi, A., Viranyi, Z., Kubinyi, E., Topal, J. & Csanyi, V. (2005). Species-specific differences and similarities in the behavior of hand-raised dog and wolf pups in social situations with humans. *Developmental Psychobiology*, *47*(2), 111-122.

Gad, S. C. (2006). *Animal models in toxicology.* CRC Press.

Gaines, S. A., Rooney, N. J. & Bradshaw, J. W. (2008). The effect of feeding enrichment upon reported working ability and behavior of kenneled working dogs. *Journal of Forensic Sciences*, *53*(6), 1400-1404.

Garner, J. (2006). Perseveration and stereotypy: systems-level insights from clinical psychology. *Stereotypic animal behaviour: fundamentals and applications to welfare*, 121e152.

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*(3), 294–304.

Giurfa, M. (2013). Cognition with few neurons: higher-order learning in insects. *Trends in Neurosciences*, *36*(5), 285–294.

Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 237-288.

Gould, S. J. (1994). *Eight little piggies: Reflections in natural history.* WW Norton & Company.

Gross, D. R. (2009). *Animal models in cardiovascular research.* New York: Springer.

Hall, L. E. (2010). *Assessing the impact of human interaction on the welfare and rehoming potential shelter dogs* (Unpublished doctoral dissertation). University of Stirling.

Hall, R. & Everds, N. (2008). Principles of clinical pathology for toxicology studies. In *Principles and methods of toxicology* (p. 1317-1358). Florida, USA: CRC Press Taylor and Francis Group.

Hamlin, R. L. (2008). How many ecg leads are required for *in vivo* studies in safety pharmacology? *Journal of Pharmacological and Toxicological Methods*, *57*(3), 161-168.

Hansen, S. W. & Jeppesen, L. L. (2006). Temperament, stereotypies and anticipatory behaviour as measures of welfare in mink. *Applied Animal Behaviour Science*, *99*(1), 172-182.

Hanton, G. & Rabemampianina, Y. (2006). The electrocardiogram of the beagle dog: reference values and effect of sex, genetic strain, body position and heart rate. *Laboratory Animals*, *40*(2), 123-136.

Harding, E. J., Paul, E. S. & Mendl, M. (2004). Animal behaviour: cognitive bias and affective state. *Nature*, *427*(6972), 312-312.

Hare, B. & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, *9*(9), 439-444.

Hartung, T. (2010). Comparative analysis of the revised directive 2010/63/eu for the protection of laboratory animals with its predecessor 86/609/eec–a t4 report. *ALTEX*, *27*(4/10), 285.

Haselton, M. G. & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47-66.

Haverbeke, A., Laporte, B., Depiereux, E., Giffroy, J.-M. & Diederich, C. (2008). Training methods of military dog handlers and their effects on the team's performances. *Applied Animal Behaviour Science*, *113*(1), 110-122.

Hawkins, P., Morton, D., Burman, O., Dennison, N., Honess, P., Jennings, M., Lane, S., Middleton, V., Roughan, J., Wells, S. & Westwood, K. (2011). A guide to defining and implementing protocols for the welfare assessment of laboratory animals: Eleventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Laboratory Animals*, *45*(1), 1-13.

Hawkins, P., Morton, D. B., Bevan, R., Heath, K., Kirkwood, J., Pearce, P., Scott, L., Whelan, G. & Webb, A. (2004). Husbandry refinements for rats, mice, dogs and non-human primates used in telemetry procedures. *Laboratory Animals*, *38*(1), 1-10.

Healy, S. D. & Tovee, M. J. (1999). Impoverishment: Neurophysiological effects. *Attitudes to Animals: Views in Animal Welfare*, 54.

Hennessy, M., Voith, V., Hawke, J., Young, T., Centrone, J., McDowell, A., Linden, F. & Davenport, G. (2002). Effects of a program of human interaction and alterations in diet composition on activity of the hypothalamic-pituitary-adrenal axis in dogs housed in a public animal shelter. *Journal of the American Veterinary Medical Association*, *221*(1), 65-71.

Hennessy, M., Voith, V., Young, T., Hawke, J., Centrone, J., McDowell, A., Linden, F. & Davenport, G. (2002). Exploring human interaction and diet effects on the

behavior of dogs in a public animal shelter. *Journal of Applied Animal Welfare Science*, *5*(4), 253-273.

Hennessy, M., Williams, M., Miller, D., Douglas, C. & Voith, V. (1998). Influence of male and female petters on plasma cortisol and behaviour: can human interaction reduce the stress of dogs in a public animal shelter? *Applied Animal Behaviour Science*, *61*(1), 63–77.

Hetts, S., Derrell Clark, J., Calpin, J., Arnold, C. & Mateo, J. (1992b). Influence of housing conditions on beagle behaviour. *Applied Animal Behaviour Science*, *34*(1-2), 137-155.

Hetts, S., Derrell Clark, J., Calpin, J. P., Arnold, C. E. & Mateo, J. M. (1992a). Influence of housing conditions on beagle behaviour. *Applied Animal Behaviour Science*, *34*(1), 137–155.

Hiby, E., Rooney, N. & Bradshaw, J. (2004). Dog training methods: Their use, effectiveness and interaction with behaviour and welfare. *Animal Welfare*, *13*(1), 63-69.

Hoffmann, M., Kastner, S. & Kramer, S. (2010). Thermal and mechanical threshold testing in beagle dogs, baseline reproducibility and the influence of sedation and opioid analgesia. In *Abstracts presented at the Association of Veterinary Anaesthetists Spring meeting, 30th-31st March 2010. Veterinary Anaesthesia and Analgesia* (Vol. 37, p. 6).

Home Office. (1986). *Animals (Scientific Procedures) Act 1986*. Her Majesty's Stationary Office.

Home Office. (1989). *Code of practice for the housing and care of animals in designated breeding and supply establishments*. Her Majesty's Stationary Office.

Home Office. (1995). *Code of practice for the housing and care of animals used in scientific procedures*. Her Majesty's Stationary Office.

Home Office. (2004). *Statistics of scientific procedures on living animals in Great Britain 2003*. Crown Copyright.

Home Office. (2005). *Statistics of scientific procedures on living animals in Great Britain 2004*. Crown Copyright.

Home Office. (2006). *Statistics of scientific procedures on living animals in Great Britain 2005*. Crown Copyright.

Home Office. (2007). *Statistics of scientific procedures on living animals in Great Britain 2006*. Crown Copyright.

Home Office. (2008). *Statistics of scientific procedures on living animals in Great Britain 2007*. Crown Copyright.

Home Office. (2009). *Statistics of scientific procedures on living animals in Great Britain 2008*. Crown Copyright.

Home Office. (2010). *Statistics of scientific procedures on living animals in Great Britain 2009* . Crown Copyright,.

Home Office. (2011). *Statistics of scientific procedures on living animals in Great Britain 2010* . Crown Copyright.

Home Office. (2012). *Statistics of scientific procedures on living animals in Great Britain 2011* . Crown Copyright.

Home Office. (2013). *Statistics of scientific procedures on living animals in Great Britain 2012* . Crown Copyright.

Home Office. (2014). *Statistics of scientific procedures on living animals in Great Britain 2013*. Crown Copyright.

Horvath, K., Angeletti, D., Nascetti, G. & Carere, C. (2013). Invertebrate welfare: an overlooked issue. *Annali dell'Istituto superiore di sanità*, *49*(1), 9-17.

Horváth, Z., Dòka, A. & Miklòsi, A. (2008). Affiliative and disciplinary behavior of human handlers during play with their dog affects cortisol concentrations in opposite directions. *Hormones and Behavior*, *54*(1), 107-114.

Hubrecht, R. (1993). A comparison of social and environmental enrichment methods for laboratory housed dogs. *Applied Animal Behaviour Science*, *37*(4), 345–361.

Hubrecht, R. (1995a). Enrichment in puppyhood and its effects on later behaviour of dogs. *Labroatory Animal Science*, *45*, 70-75.

Hubrecht, R. (1995b). *The welfare of dogs in human care* (J. Serpell, Ed.). Cambridge: Cambridge University Press.

Hubrecht, R. (2010). Enrichment: Animal welfare and experimental outcomes. In R. Hubrecht & J. Kirkwell (Eds.), *The UFAW handbook on the care and management of laboratory and other research animals: eighth edition.*

Hubrecht, R. & Kirkwood, J. (2010). Introduction. In R. Hubrecht & J. Kirkwell (Eds.), (p. 1-2).

Hubrecht, R. & Serpell, J. (1993). Influence of housing conditions on the behaviour and welfare of dogs. *Applied Animal Behavior Science*, *35*.

Hubrecht, R., Serpell, J. & Poole, T. (1992). Correlates of pen size and housing conditions on the behaviour of kennelled dogs. *Applied Animal Behaviour Science*, *34*(4), 365–383.

Hunt, J., Grint, N., Murrell, J. & Taylor, P. (2010). Prospective, randomized, blinded comparison of the sedative and analgesic effects of premedication with dexmedetomidine/buprenorphine and acepromazine/buprenorphine in dogs and cats-interim results. In *Spring meeting of the association of veterinary anaesthetists* (p. 22).

International Conference on Harmonization. (2008). *Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals.* Retrieved from www.ich.org

International Conference on Harmonization. (2009). *ICH Topic M 3(R2) Non-Clinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals.* Retrieved from www.ich.org

Ipsos MORI. (2012). *Views on the use of animals in scientific research.* London: Department for Business, Innovatio and Skill.

Jolliffe, C., Flack, D., Joyce, B., Keeley, J., Taylor, P. & Watts, G. (2009). Mechanical nociceptive threshold testing in assessing pain. *The Veterinary Record*, *165*, 122.

Jones, A. (2007). Sensory development in puppies (*Canis lupus f. familiaris*): Implications for improving canine welfare. *Animal Welfare*, 319-329.

Keen, H. A., Nelson, O. L., Robbins, C. T., Evans, M., Shepherdson, D. J. & Newberry, R. C. (2013). Validation of a novel cognitive bias task based on difference in quantity of reinforcement for assessing environmental enrichment. *Animal Cognition*, 1-13.

Keller, S. E., Weiss, J. M., Schleifer, S. J., Miller, N. E. & Stein, M. (1981). Suppression of immunity by stress: effect of a graded series of stressors on lymphocyte stimulation in the rat. *Science*, *213*(4514), 1397-1400.

Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M. & Altman, D. G. (2010). Improving bioscience research reporting: the arrive guidelines for reporting animal research. *PLoS Biology*, *8*(6), e1000412.

Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F., Cuthill, I. C., Fry, D., Hutton, J. & Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One*, *4*(11), e7824.

Klauenberg, S., Maier, C., Assion, H.-J., Hoffmann, A., Krumova, E. K., Magerl, W., Scherens, A., Treede, R.-D. & Juckel, G. (2008). Depression and changed pain perception: hints for a central disinhibition mechanism. *Pain*, *140*(2), 332-343.

Koerner, J. E. & Siegl, P. K. (2013). Safety pharmacology: guidelines S7A and S7B. In *Global approach in safety testing* (p. 243-265). Springer.

König, H. E., Liebich, H.-G. et al. (2004). *Veterinary anatomy of domestic mammals: textbook and colour atlas.* Schattauer GmbH.

Koolhaas, J., Korte, S., De Boer, S., Van Der Vegt, B., Van Reenen, C., Hopster, H., De Jong, I., Ruis, M. & Blokhuis, H. (1999). Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*(7), 925-935.

Kraeuter, K. (2001). *Training your beagle.* U.S.A.: Barron's Educational Series.

Kubinyi, E., Viranyi, Z. & Miklosi, A. (2007). Comparative social cognition: From wolf and dog to humans. *Comparative Cognition & Behavior Reviews*, *2*, 26-46.

Langer, A. W., Obrist, P. A. & McCubbin, J. A. (1979). Hemodynamic and metabolic adjustments during exercise and shock avoidance in dogs. *American Journal of*

Physiology-Heart and Circulatory Physiology, *236*(2), H225–H230.

Latham, N. (2010). Brief introduction to welfare assessment: A 'toolbox' of techniques. In R. Hubrecht & J. Kirkwell (Eds.), *The UFAW handbook on the care and management of laboratory and other research animals: eighth edition* (p. 76-91).

Laule, G. (2010). Positive reinforcement training for laboratory animals. In R. Hubrecht & J. Kirkwell (Eds.), *The ufaw handbook on the care and management of laboratory and other research animals: Eighth edition* (p. 206-218). Wheathampstead, UK: Universities Federation for Animal Welfare.

Leishman, D., Beck, T., Dybdal, N., Gallacher, D., Guth, B., Holbrook, M., Roche, B. & Wallis, R. (2012). Best practice in the conduct of key nonclinical cardiovascular assessments in drug development: Current recommendations from the safety pharmacology society. *Journal of Pharmacological and Toxicological Methods*, *65*(3), 93-101.

Lewington, S., Clarke, R., Qizilbash, N., Peto, R. & Collins, R. (2002). Prospective studies collaboration. age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*, *360*(9349), 1903-1913.

Lyons, D. M., Parker, K. J., Katz, M. & Schatzberg, A. F. (2009). Developmental cascades linking stress inoculation, arousal regulation, and resilience. *Frontiers in Behavioral Neuroscience*, *3*, 32.

MacDougall, J., Tuxen, D., Sale, D. & Moroz, J. (1985). Arterial blood pressure response to heavy exercise. *Journal of Applied Physiology*, *58*, 785-790.

MacMahon, S., Peto, R., Collins, R., Godwin, J., MacMahon, S., Cutler, J., Sorlie, P., Abbott, R., Collins, R., Neaton, J. et al. (1990). Blood pressure, stroke, and coronary heart disease: part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *The Lancet*, *335*(8692), 765–774.

Marsh, J. T., Lavender, J. F., Chang, S.-S. & Rasmussen, A. (1963). Poliomyelitis in monkeys: decreased susceptibility after avoidance stress. *Science*, *140*(3574), 1414-1415.

Mason, G. & Latham, N. (2004). Can't stop, won't stop: Is stereotypy a reliable animal welfare indicator? *Animal Welfare*, *13*(SUPPL.), S57-S69.

Matheson, S., Asher, L. & Bateson, M. (2008). Larger, enriched cages are associated with 'optimistic' response biases in captive european starlings (*Sturnus vulgaris*. *Applied Animal Behaviour Science*, *109*, 374-383.

McKinley, J., Buchanan-Smith, H. M., Bassett, L. & Morris, K. (2003). Training common marmosets (callithrix jacchus) to cooperate during routine laboratory procedures: Ease of training and time investment. *Journal of Applied Animal Welfare Science*, *6*(3), 209-220.

Meaney, M. J., Aitken, D. H., Van Berkel, C., Bhatnagar, S. & Sapolsky, R. M. (1988). Effect of neonatal handling on age-related impairments associated with the hippocampus. *Science*, *239*(4841), 766-768.

Meaney, M. J., O'Donnell, D., Rowe, W., Tannenbaum, B., Steverman, A., Walker, M., Nair, N. & Lupien, S. (1995). Individual differences in hypothalamic-pituitary-adrenal activity in later life and hippocampal aging. *Experimental Gerontology*, *30*(3), 229-251.

Mech, L. D. & Boitani, L. (2003). *Wolves: behavior, ecology, and conservation.* Chicago: University of Chicago Press.

Meek, P. D. (1999). The movement, roaming behaviour and home range of free-roaming domestic dogs, *Canis lupus familiaris*, in coastal New South Wales. *Wildlife Research*, *26*(6), 847–855.

Meijer, A. J. & Codogno, P. (2006). Signalling and autophagy regulation in health, aging and disease. *Molecular Aspects of Medicine*, *27*(5), 411-425.

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E. & Casey, R. (2010). Dogs showing separation-related behaviour exhibit a 'pessimistic' cognitive bias. *Current Biology*, *20*(19), R839-R840.

Mendl, M., Burman, O., Parker, R. & Paul, E. (2009). Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, *118*(3-4), 161-181.

Mendl, M., Burman, O. & Paul, E. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2895-2904.

Mendl, M., Paul, E. S. & Chittka, L. (2011). Animal behaviour: Emotion in invertebrates? *Current Biology*, *21*(12), R463–R465.

Mertens, P. A. & Unshelm, J. (1996). Effects of group and individual housing on the behavior of kennelled dogs in animal shelters. *Anthrozoos*, *9*(1), 40–51.

Meunier, L. D. (2006). Selection, acclimation, training, and preparation of dogs for the research setting. *ILAR Journal*, *47*(4), 326-347.

Mitchell, C. B. T. S. R., A.Z. McMahon. (2010). Sensitivity of two noninvasive blood pressure measurement techniques compared to telemetry in cynomolgus monkeys and beagle dogs. *Journal of Pharmacological and Toxicological Methods*, *62*, 64-63.

Moodie, E. M. & Chamove, A. S. (1990). Brief threatening events beneficial for captive tamarins? *Zoo Biology*, *9*(4), 275-286.

Moscardo, E., Fasdelli, N., Giarola, A., Tontodonati, M. & Dorigatti, R. (2009). An optimised neurobehavioural observation battery integrated with the assessment of cardiovascular function in the beagle dog. *Journal of Pharmacological and Toxicological Methods*, *60*(2), 198-209.

Müller, C. A., Riemer, S., Rosam, C. M., Schößwender, J., Range, F. & Huber, L. (2012). Brief owner absence does not induce negative judgement bias in pet dogs. *Animal Cognition*, *15*(5), 1031–1035.

National Animal Ethics Advisory Committee, Ministry for Primary Industries New Zealand. (2013). *Annual report 1 January to 31 December 2012*. NAEAC.

National Centre for the 3Rs. (2010). *The challenge of animal research*. Retrieved from http://secure.nc3rs.org.uk/document.asp?id=1304

Neave, H. W., Daros, R. R., Costa, J. H., von Keyserlingk, M. A. & Weary, D. M. (2013). Pain and pessimism: Dairy calves exhibit negative judgement bias following hot-iron disbudding. *PloS one*, *8*(12), e80556.

Nerem, R. M., Levesque, M. J. & Cornhill, J. F. (1980). Social environment as a factor in diet-induced atherosclerosis. *Science*, *208*(4451), 1475-1476.

Normando, S., Corain, L., Salvadoretti, M., Meers, L. & Valsecchi, P. (2009). Effects of an enhanced human interaction program on shelter dogs' behaviour analysed using a novel nonparametric test. *Applied Animal Behaviour Science*, *116*(2-4), 211-219.

Normando, S., Meers, L., Simontacchi, C., Coultis, D., Odberg, F. & Bono, G. (2005). Does interruption in a 'walking program' cause stress in kennelled beagles? In *Proceeding of the 14th annual conference of the international society for anthrozoology* (p. 51-52).

Nuffield Council on Bioethics. (2005). *The ethics of research involving animals*. London, UK: Nuffield Council on Bioethics.

Odendaal, J. & Meintjes, R. (2003). Neurophysiological correlates of affiliative behaviour between humans and dogs. *The Veterinary Journal*, *165*(3), 296–301.

Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., Lilly, P., Sanders, J., Sipes, G., Bracken, W. et al. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory Toxicology and Pharmacology*, *32*(1), 56-67.

Overmier, J., Patterson, J. & Wielkiewicz, R. (1980). Environmental contingencies as sources of stress in animals. In S. Levine & H. Ursin (Eds.), *Coping and health*. New York: Plenum Press.

Palmer, R. & Custance, D. (2008). A counterbalanced version of ainsworth's strange situation procedure reveals secure-base effects in dog-human relationships. *Applied Animal Behaviour Science*, *109*(2), 306-319.

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neuroscience & Biobehavioral Reviews*, *35*(9), 1791-1804.

Paul, E. S., Harding, E. J. & Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience & Biobehavioral Reviews*, *29*(3), 469-491.

Peers, I. S., South, M. C., Ceuppens, P. R., Bright, J. D. & Pilling, E. (2014). Can you trust your animal study data? *Nature Reviews Drug Discovery*, *13*, 560.

Pellegatti, M. (2013). Dogs and monkeys in preclinical drug development: The challenge of reducing and replacing. *Expert Opinion on Drug Metabolism and Toxicology*, *9*(9), 1171-1180.

Phillips, B., Smith, D., Combes, R., G. Descotes, S. J., Hack, R., Kemkowski, J., Krauser, K., Pfister, R., Rabemampianina, Y., Sparrow, S., Stephan-Gueldner, M. & von Landenberg, F. (2004). An approach to minimise dog use in regulatory toxicology: Production of a best practice guide to study design. *ALTA*, *33*(Supplement 1), 447-451.

Pincus, T. & Morley, S. (2001). Cognitive-processing bias in chronic pain: a review and integration. *Psychological Bulletin*, *127*(5), 599.

Polson, S., Taylor, P. M. & Yates, D. (2012). Analgesia after feline ovariohysterectomy under midazolam-medetomidine-ketamine anaesthesia with buprenorphine or butorphanol, and carprofen or meloxicam: a prospective, randomised clinical trial. *Journal of Feline Medicine and Surgery*, *14*(8), 553-559.

Pomerantz, O., Terkel, J., Suomi, S. & Paukner, A. (2012). Stereotypic head twirls, but not pacing, are related to a 'pessimistic'-like judgment bias among captive tufted capuchins (*Cebus apella*). *Animal Cognition*, *15*(4), 689-698.

Poole, T. (1992). The nature and evolution of behavioural needs in mammals. *Animal Welfare*, *1*(3), 203–220.

Poole, T. (1997). Happy animals make good science. *Laboratory Animals*, *31*(2), 116-124.

Prescott, M. (2010). Ethics of primate use. *Advances in Science and Research*, *5*(1), 11-22.

Prescott, M. & Buchanan-Smith, H. M. (2007). Training laboratory-housed non-human primates, part 1: a UK survey. *Animal Welfare*, *16*(1), 21-36.

Prescott, M., Morton, D. B., Anderson, D., Buckwell, A., Heath, S., Hubrecht, R., Jennings, M., Robb, D., Ruane, B., Swallow, J. & Thompson, P. (2004). Refining dog husbandry and care: Eighth report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Laboratory Animals*, *38*(SUPPL. 1), S1:1-S1:94.

Pretterer, G., Bubna-Littitz, H., Windischbauer, G., Gabler, C. & Griebel, U. (2004). Brightness discrimination in the dog. *Journal of Vision*, *4*(3), 10.

Prisant, L., Carr, A. & Hawkins, D. (1993). Treating hypertensive emergencies. controlled reduction of blood pressure and protection of target organs. *Postgraduate Medicine*, *93*(2), 92-6.

Pruessner, J. C., Hellhammer, D. H. & Kirschbaum, C. (1999). Burnout, perceived

stress, and cortisol responses to awakening. *Psychosomatic Medicine*, *61*(2), 197–204.

Pryce, C. R., Ruedi-Bettschen, D., Dettling, A. C., Weston, A., Russig, H., Ferger, B. & Feldon, J. (2005). Long-term effects of early-life environmental manipulations in rodents and primates: potential animal models in depression research. *Neuroscience & Biobehavioral Reviews*, *29*(4), 649–674.

Rampon, C., Tang, Y.-P., Goodhouse, J., Shimizu, E., Kyin, M. & Tsien, J. Z. (2000). Enrichment induces structural changes and recovery from nonspatial memory deficits in ca1 nmdar1-knockout mice. *Nature Neuroscience*, *3*(3), 238-244.

Reite, M. (1987). Infant abuse and neglect: lessons from the primate laboratory. *Child Abuse & Neglect*, *11*(3), 347-355.

Richmond, J. (2010). The Three Rs. In R. Hubrecht & J. Kirkwell (Eds.), *The UFAW Handbook on the Care and Management of Laboratory and Other Research Animals: Eighth Edition* (p. 3-22).

Richter, S. H., Garner, J. P., Auer, C., Kunert, J. & Wurbel, H. (2010). Systematic variation improves reproducibility of animal experiments. *Nature Methods*, *7*(3), 167-168.

Robinson, S., Chapman, K., Hudson, S., S, S., Spencer-Briggs, D., Danks, A., Hill, R., Everett, D., Mulier, B., Old, S. & Bruce, C. (2009). *Guidance on dose level selection for regulatory general toxicology studies for pharmaceuticals.* London, UK: NC3Rs/LASA.

Robinson, S., Delongeas, J.-L., Donald, E., Dreher, D., Festag, M., Kervyn, S., Lampo, A., Nahas, K., Nogues, V., Ockert, D. et al. (2008). A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. *Regulatory Toxicology and Pharmacology*, *50*(3), 345-352.

Robinson, V. (2005). Finding alternatives: an overview of the 3Rs and the use of animals in research. *School Science Review*, *87*, 111-114.

Rogers, L. J. (2008). Development and function of lateralization in the avian brain. *Brain Research Bulletin*, *76*(3), 235–244.

Rogers, L. J. (2010). Relevance of brain and behavioural lateralization to animal welfare. *Applied Animal Behaviour Science*, *127*(1), 1–11.

Rogers, L. J., Rigosi, E., Frasnelli, E. & Vallortigara, G. (2013). A right antenna for social behaviour in honeybees. *Scientific Reports*, *3*, 2045.

Rogers, L. J. & Vallortigara, G. (2008). From antenna to antenna: lateral shift of olfactory memory recall by honeybees. *PLoS One*, *3*(6), e2340.

Rosado, B., García-Belenguer, S., León, M., Chacón, G., Villegas, A. & Palacio, J. (2010). Blood concentrations of serotonin, cortisol and dehydroepiandrosterone in aggressive dogs. *Applied Animal Behaviour Science*, *123*(3), 124–130.

Rushen, J., Taylor, A. A. & de Passillé, A. M. (1999). Domestic animals' fear of humans and its effect on their welfare. *Applied Animal Behaviour Science*, *65*(3), 285-303.

Russell, W. & Burch, R. (1959). *The principles of humane experimental technique.* Wheathampstead, UK: Universities Federation for Animal Welfare.

Ruys, J., Mendoza, S., Capitanio, J. & Mason, W. (2004). Behavioral and physiological adaptation to repeated chair restraint in rhesus macaques. *Physiology and Behavior*, *82*(2-3), 205-213.

Rygula, R., Pluta, H. & Popik, P. (2012). Laughing rats are optimistic. *PloS One*, *7*(12), e51959.

Sales, G., Hubrecht, R., Peyvandi, A., Milligan, S. & Shield, B. (1997). Noise in dog kennelling: is barking a welfare problem for dogs? *Applied Animal Behaviour Science*, *52*(3), 321–329.

Salmeto, A. L., Hymel, K. A., Carpenter, E. C., Brilot, B. O., Bateson, M. & Sufka, K. J. (2011). Cognitive bias in the chick anxiety–depression model. *Brain Research*, *1373*, 124-130.

Sambrook, T. & Buchanan-Smith, H. (1997). Control and complexity in novel object enrichment. *Animal Welfare*, *6*(3), 207-216.

Schipper, L. L., Vinke, C. M., Schilder, M. B. & Spruijt, B. M. (2008). The effect of feeding enrichment toys on the behaviour of kennelled dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, *114*(1), 182-195.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433-440.

Scott, J. P. (1965). *Genetics and the social behavior of the dog.* University of Chicago Press.

Scott, J. P. & Fuller, J. L. (1974). *Dog behavior.* University Press.

Seabrook, M. F. (1984). The psychological interaction between the stockman and his animals and its influence on performance of pigs and dairy cows. *Veterinary Record*, *115*, 84-87.

Seery, M. D. (2011). Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience & Biobehavioral Reviews*, *35*(7), 1603–1610.

Seery, M. D., Holman, E. A. & Silver, R. C. (2010). Whatever does not kill us: cumulative lifetime adversity, vulnerability, and resilience. *Journal of Personality and Social Psychology*, *99*(6), 1025.

Selye, H. (1946). The general adaptation syndrome and the diseases of adaptation 1. *The Journal of Clinical Endocrinology & Metabolism*, *6*(2), 117-230.

Serpell, J. (1995). *The domestic dog: its evolution, behaviour and interactions with people.* Cambridge University Press.

Shaw, R., Festing, M., Peers, I. & Furlong, L. (2002). Use of factorial designs to optimize animal experiments and reduce animal use. *ILAR Journal*, *43*(4), 223–232.

Shi, M., Qi, W.-J., Gao, G., Wang, J.-Y. & Luo, F. (2010). Increased thermal and mechanical nociceptive thresholds in rats with depressive-like behaviors. *Brain Research*, *1353*, 225-233.

Siniscalchi, M., Quaranta, A. & Rogers, L. J. (2008). Hemispheric specialization in dogs for processing different acoustic stimuli. *PLoS One*, *3*(10), e3349.

Siniscalchi, M., Sasso, R., Pepe, A. M., Dimatteo, S., Vallortigara, G. & Quaranta, A. (2011). Sniffing with the right nostril: lateralization of response to odour stimuli by dogs. *Animal Behaviour*, *82*(2), 399–404.

Siniscalchi, M., Sasso, R., Pepe, A. M., Vallortigara, G. & Quaranta, A. (2010). Dogs turn left to emotional stimuli. *Behavioural Brain Research*, *208*(2), 516–521.

Sjaastad, O. V., Hove, K. & Sand, O. (2010). *Physiology of domestic animals*. Oslo, Norway: Scandinavian Veterinary Press.

Slaughter, M., Birmingham, J., Patel, B., Whelan, G., Krebs-Brown, A., Hockings, P. & Osborne, J. (2002). Extended acclimatization is required to eliminate stress effects of periodic blood-sampling procedures on vasoactive hormones and blood volume in beagle dogs. *Laboratory Animals*, *36*(4), 403-410.

Smith, D., Broadhead, C., Descotes, G., Fosse, R., Hack, R., Krauser, K., Phillips, B., Rabemampianina, Y., Sanders, J., Sparrow, S. et al. (2002). Preclinical safety evaluation using nonrodent species: an industry/welfare project to minimize dog use. *ILAR Journal*, *43*(Suppl 1), S39-S42.

Spotte, S. (2012). *Societies of wolves and free-ranging dogs*. Cambridge University Press.

Starkey, M. P., Scase, T. J., Mellersh, C. S. & Murphy, S. (2005). Dogs really are man's best friend - canine genomics has applications in veterinary and human medicine! *Briefings in Functional Genomics & Proteomics*, *4*(2), 112-128.

Stephens, M. L. (2011). Pursuing medawar's challenge for full replacement. *Alternatives to Animal Use in the Life Sciences, Montreal 2011*, 23.

Svartberg, K. & Forkman, B. (2002). Personality traits in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, *79*(2), 133-155.

Tasker, L. (2012). *Linking welfare and quality of scientific output in cynomolgus macaques (Macaca fascicularis) used for regulatory toxicology* (Unpublished doctoral dissertation). University of Stirling, Stirling, UK.

Tattersall, M. L., Dymond, M., Hammond, T. & Valentin, J.-P. (2006). Correction of QT values to allow for increases in heart rate in conscious Beagle dogs in toxicology assessment. *Journal of Pharmacological and Toxicological Methods*, *53*(1), 11–19.

Taylor, K., Gordon, N., Langley, G. & Higgins, W. (2008). Estimates for worldwide laboratory animal use in 2005. *ATLA*, *36*(3), 327.

Taylor, K. & Mills, D. (2007a). The effect of the kennel environment on canine welfare: A critical review of experimental studies. *Animal Welfare*, *16*(4), 435-447.

Taylor, K. & Mills, D. (2007b). Is quality of life a useful concept for companion animals? *Animal Welfare*, *16*(Supplement 1), 55-65.

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V., Sawyer, S., Greenfield, D., Germonpré, M., Sablin, M., López-Giráldez, F., Domingo-Roura, X. et al. (2013). Complete mitochondrial genomes of ancient canids suggest a european origin of domestic dogs. *Science*, *342*(6160), 871–874.

The Kennel Club. (2011). *The Kennel Club's illustrated breed standards*. Great Britain: Ebury Publishing.

Turner, A. M. & Greenough, W. T. (1985). Differential rearing effects on rat visual cortex synapses. i. synaptic and neuronal density and synapses per neuron. *Brain Research*, *329*(1), 195–203.

United States Department of Agriculture. (2013). *Animal Welfare Act*. Washington: USDA.

USDA. (2014). *Annual report animal usage by fiscal year 2012*. United States Department of Agriculture.

Valentin, J.-P. (2010). Reducing qt liability and proarrhythmic risk in drug discovery and development. *British Journal of Pharmacology*, *159*(1), 5-11.

Van Citters, R. L. & Franklin, D. L. (1969). Cardiovascular performance of Alaska sled dogs during exercise. *Circulation Research*, *24*(1), 33-42.

van der Harst, J. E., Baars, A.-M. & Spruijt, B. M. (2003). Standard housed rats are more sensitive to rewards than enriched housed rats as reflected by their anticipatory behaviour. *Behavioural Brain Research*, *142*(1), 151-156.

Van Praag, H., Kempermann, G. & Gage, F. H. (2000). Neural consequences of enviromental enrichment. *Nature Reviews Neuroscience*, *1*(3), 191–198.

Vatner, S. F., Higgins, C. B., White, S., Patrick, T. & Franklin, D. (1971). The peripheral vascular response to severe exercise in untethered dogs before and after complete heart block. *Journal of Clinical Investigation*, *50*(9), 1950.

Verbeek, E., Ferguson, D. & Lee, C. (2014). Are hungry sheep more pessimistic? the effects of food restriction on cognitive bias and the involvement of ghrelin in its regulation. *Physiology and Behavior*, *123*, 67-75.

Vilà, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., Crandall, K. A., Lundeberg, J. & Wayne, R. K. (1997). Multiple and ancient origins of the domestic dog. *Science*, *276*(5319), 1687–1689.

Villemure, C. & Bushnell, M. C. (2002). Cognitive modulation of pain: how do attention and emotion influence pain processing? *Pain*, *95*(3), 195-199.

Volkmar, F. R. & Greenough, W. T. (1972). Rearing complexity affects branching of dendrites in the visual cortex of the rat. *Science*, *176*(4042), 1445-1447.

Waiblinger, S., Menke, C., Korff, J. & Bucher, A. (2004). Previous handling and gentle interactions affect behaviour and heart rate of dairy cows during a veterinary procedure. *Applied Animal Behaviour Science*, *85*(1), 31-42.

Wallace, J., Sanford, J., Smith, M. & Spencer, K. (1990). The assessment and control of the severity of scientific procedures on laboratory animals: Report of the Laboratory Animal Science Sssociation working party (Assessment and control of severity). *Laboratory Animals*, *24*(2), 97-130.

Wells, D. (2004a). The influence of toys on the behaviour and welfare of kennelled dogs. *Animal Welfare*, *13*(3), 367-373.

Wells, D. (2004b). A review of environmental enrichment for kennelled dogs, canis familiaris. *Applied Animal Behaviour Science*, *85*(3-4), 307-317.

Welp, T., Rushen, J., Kramer, D., Festa-Bianchet, M. & De Passille, A. (2004). Vigilance as a measure of fear in dairy cattle. *Applied Animal Behaviour Science*, *87*(1), 1-13.

Wiepkema, P. & Koolhaas, J. (1993). Stress and animal welfare. *Animal Welfare*, *2*(3), 195-218.

Wilson, H. D., Boyette-Davis, J. & Fuchs, P. N. (2007). The relationship between basal level of anxiety and the affective response to inflammation. *Physiology and Behavior*, *90*(2), 506–511.

Wilsson, E. & Sundgren, P.-E. (1998). Behaviour test for eight-week old puppiesheritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*, *58*(1), 151–162.

Wolfensohn, S. & Honess, P. (2008). *Handbook of primate husbandry and welfare.* London, UK: John Wiley & Sons.

Wolfer, D. P., Litvin, O., Morf, S., Nitsch, R. M., Lipp, H.-P. & Würbel, H. (2004). Laboratory animal welfare: cage enrichment and mouse behaviour. *Nature*, *432*(7019), 821–822.

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, *26*(3), 263–263.

Würbel, H. (2001). Ideal homes? Housing effects on rodent brain and behaviour. *Trends in Neurosciences*, *24*(4), 207–211.

Würbel, H. (2002). Behavioral phenotyping enhanced - beyond (environmental) standardization. *Genes, Brain and Behavior*, *1*(1), 3-8.

Würbel, H. & Garner, J. P. (2007). Refinement of rodent research through environmental enrichment and systematic randomization. *NC3Rs*, *9*, 1-9.

Yagami, K., Mashimo, T., Sekiguchi, F., Sugiyama, F., Yamamura, K. & Serikawa, T. (2010). Survey of live laboratory animals reared in japan (2009). *Experimental Animals*, *59*.

Yeates, J. & Main, D. (2008). Assessment of positive welfare: a review. *The Veterinary Journal*, *175*(3), 293-300.

Youatt, W. (1852). *The Dog.* UK: Blanchard and Lea.

Zampaglione, B., Pascale, C., Marchisio, M. & Cavallo-Perin, P. (1996). Hypertensive urgencies and emergencies. prevalence and clinical presentation. *Hypertension*, *27*, 144-147.

Ziegler, T. E., Scheffler, G. & Snowdon, C. T. (1995). The relationship of cortisol levels to social environment and reproductive functioning in female cotton-top tamarins, saguinus oedipus. *Hormones and Behavior*, *29*(3), 407-424.

Zurlo, J., Bayne, K., Cimino Brown, D., Burkholder, T., Dellarco, V., Ellis, A., Garrett, L., Hubrecht, R., Janus, E., Kinter, R., Luddy, E., Meunier, L., Scorpio, D., Serpell, J. & Todhunter, R. (2011). Critical evaluation of the use of dogs in biomedical research and testing. *ALTEX*, *4*, 355.