

**Improving Captive Animal Welfare Through the Application of Cognitive
Enrichment**

**Submitted by Louise Natalie Millar to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Psychology
In August 2013**

**This thesis is available for Library use on the understanding that it is
copyright material and that no quotation from the thesis may be published
without proper acknowledgement.**

**I certify that all material in this thesis which is not my own work has been
identified and that no material has previously been submitted and
approved for the award of a degree by this or any other University.**

Signature:

Abstract

Welfare for captive animals in laboratory, farm, companion and zoo settings is difficult to assess and subject to interpretation. The perceived needs of animals change over time and according to budget, fashion and policy. The assessment of the value of environmental enrichment for captive animals is hampered by the lack of consistent, quantitative, objective and methodical research. Enrichment devices and other welfare improvement strategies lack long-term assessment or implementation. The concept of cognitive enrichment, or enabling captive animals to use their cognitive abilities to solve problems and meet challenges in their environment, is in the initial stages of development. Preliminary findings indicate that cognitive enrichment seems to be an effective method of improving captive animal welfare.

In the research described in this thesis, several welfare measurement techniques including changes in activity budget, stereotypic behaviour and cognitive bias were compared and used to measure the effects of cognitive enrichment upon captive pigeons and dogs in order to obtain a useable cognitive enrichment paradigm that can be extrapolated to many types of captive animals.

Key findings were that cognitive enrichment improved captive group-housed pigeon welfare and individually-kennelled dog welfare, indicated by significant changes in activity budget, increases in 'optimism', and reduction of stereotypic behaviour in subjects. Subjects used both low-tech and high-tech cognitive enrichments as intended and showed no sign of habituation. It is hoped that these findings will be used to improve captive animal welfare, and that the cognitive enrichment and cognitive bias paradigms developed will add to this field of research.

Table of Contents

Abstract	2
1.1 Acknowledgements	8
1.2 Declarations	9
1.3 Ethics	10
2.0 Chapter One	12
2.1 Literature review and general introduction: Does cognitive enrichment improve captive animal welfare?	12
Abstract	12
2.2 Introduction	14
2.2.1 The Welfare of Captive Animals	14
<i>Developing a Concept of Animal Welfare and Animal Rights</i>	14
<i>How can Animal Welfare be Defined?</i>	14
<i>Responsibilities of Care-Givers</i>	15
2.2.2 Environmental Enrichment	16
<i>What is Environmental Enrichment?</i>	16
<i>How Does Environmental Enrichment Affect Animal Welfare?</i>	16
<i>EE and Intrinsic Behavioural Needs</i>	18
<i>Current Challenges in Environmental Enrichment Research and Implementation</i>	18
2.2.3 Types of Environmental Enrichment	20
<i>Social Enrichment – A Brief Review</i>	20
<i>Sensory Enrichment – A Brief Review</i>	22
<i>Physical Enrichment – A Brief Review</i>	23
<i>Nutritional Enrichment – A Brief Review</i>	24
2.2.4 Cognitive Enrichment – An In-Depth Review	24
2.3.1 Measuring the Welfare Benefits of Environmental Enrichment: Why Measure Welfare?	37
2.3.2 <i>How Should Welfare be Measured?</i>	37
2.3.3 <i>Useful Welfare Measures</i>	38
<i>Behaviour</i>	38
<i>Physical Condition including Glucocorticoids</i>	40
<i>Stereotypic Behaviour</i>	40
2.3.4 <i>The Link Between Operant Response and Emotion</i>	44
3.0 Contribution of this Thesis	55
<i>Questions to be Addressed by this Thesis</i>	56
4.0 Introduction to Experiments One and Two	57

4.1 Chapter Two: Experiment One and Two	59
A complex cognitive enrichment procedure improves the welfare of group-housed pigeons to a greater extent than a physical enrichment procedure, and group presentation is preferred to individual enrichment. 59	
Abstract	59
4.2 Introduction	61
4.2.1 Environmental Enrichment (EE)	61
4.2.2 Cognitive Enrichment (CE).....	61
4.2.3 Measuring Animal Welfare	63
4.2.4 Cognitive Bias as a Welfare Measurement Tool in Non-Human Animals.....	64
4.2.5 Task Complexity.....	65
4.2.6 Group and Individual Presentation.....	65
4.2.7 Aims and Hypotheses	67
Experiment One.....	67
Experiment Two.....	68
4.3 Method	69
4.3.1 Design (Expt. 1).....	69
4.3.2 Design (Expt. 2).....	70
4.3.3 Subjects and Housing (Expt. 1 and 2)	70
4.3.4 Reliability Assessment (Expt. 1 and 2)	71
4.3.5 Activity Budget Measurement (Expt. 1)	72
4.3.6 Physical Condition Measurement (Expt. 1).....	72
4.3.7 Stereotypic Behaviour Measurement (Expt. 1 and 2)	73
4.3.8 Cognitive Bias Task Training and Probe Trials (Expt. 1 and 2).....	74
4.3.9 Dominance Observations (Expt. 2)	77
4.3.10 Cognitive Enrichment (CE) Condition (Expt. 1).....	78
4.3.11 Physical Enrichment (PE) Condition (Expt. 1).....	79
4.3.12 Cognitive Enrichment (CE) Condition (Expt. 2).....	80
4.4 Results	81
4.4.1 Activity Budget (Expt. 1).....	82
4.4.2 Physical condition (Expt. 1)	84
4.4.3 Stereotypic behaviour (Expt. 1)	85
4.4.4 Optimism (Expt. 1)	87
4.4.5 Use of Enrichment (Expt. 1)	90
4.4.6 Predicting Welfare Improvement (Expt. 1)	91
4.4.7 Stereotypic behaviour (Expt. 2)	92
4.4.8 Optimism (Expt. 2)	93
4.4.9 Use of Enrichment (Expt. 2)	96

4.4.10 Group versus Individual Enrichment Presentation (Expt. 2) ..	98
4.4.11 Dominance (Expt. 2).....	100
4.4.12 Predicting Welfare Improvement (Expt. 2).....	103
4.5 Discussion	104
4.5.1 Activity Budget (Expt. 1).....	104
4.5.2 Physical Condition (Expt. 1)	105
4.5.3 Stereotypic Behaviour (Expt. 1).....	105
4.5.4 Optimism (Expt. 1)	107
4.5.5 Use of Enrichment (Expt. 1)	109
4.5.6 Predicting Welfare Improvement (Expt. 1).....	110
4.5.7 Order Effects (Expt. 1).....	111
4.5.8 Stereotypic behaviour (Expt. 2)	111
4.5.9 Optimism (Expt. 2)	112
4.5.10 Use of Enrichment (Expt. 2).....	113
4.5.11 Group versus Individual Enrichment Presentation (Expt. 2)	114
4.5.12 Dominance (Expt. 2).....	117
4.5.13 Predicting Welfare Improvement (Expt. 2).....	118
4.5.14 Conclusions (Expt. 1)	118
4.5.15 Conclusions (Expt. 2)	119
5.0 Chapter Three.....	120
The design and implementation of a cognitive enrichment device known as the DogBox	120
Abstract	120
5.1 Introduction	121
5.2 DogBox Development	122
5.3 DogBox Design	123
5.4 Programming.....	125
5.5 Key Benefits	129
<i>Effective at providing CE to captive animals</i>	129
<i>Adaptable and Flexible.....</i>	129
<i>Multi-purpose.....</i>	129
<i>Accessible</i>	130
<i>Useful Research Tool.....</i>	130
<i>Safe</i>	130
<i>Portable</i>	131
<i>Quiet and unobtrusive</i>	131
5.6 Conclusions.....	131
6.0 Introduction to Experiments Three and Four	132

6.1 Chapter Four: Experiment Three and Four	133
In kennelled dogs (<i>Canis familiaris</i>), stereotypic behaviour and learning ability vary according to length of stay, and cognitive enrichment improves welfare.....	133
Abstract	133
6.2 Introduction	135
6.2.1 <i>Dogs in Isolation.....</i>	135
6.2.2 <i>Can Enrichment Improve Kennelled Dog Welfare?.....</i>	137
6.2.3 <i>Cognitive Bias Measurement and the Advance Key Procedure.....</i>	138
6.2.4 <i>Aims and Hypotheses</i>	139
<i>Experiment Three.....</i>	139
<i>Experiment Four.....</i>	140
6.3 Method	140
6.3.1 <i>Design (Expt. 3).....</i>	140
6.3.2 <i>Design (Expt. 4).....</i>	141
6.3.3 <i>Housing.....</i>	142
6.3.4 <i>Subjects</i>	143
6.3.5 <i>Reliability Assessment.....</i>	144
6.3.6 <i>Activity Budget Observations (Expt. 3 and 4).....</i>	144
6.3.7 <i>Cognitive Bias Task (Expt. 3 and 4).....</i>	145
<i>Advance Key Procedure (Expt. 4)</i>	147
<i>Probe Trials (Expt. 4).....</i>	147
6.4 Results	148
6.4.1 <i>Site, Length of Stay and Behaviour (Expt. 3)</i>	148
6.4.2 <i>Effect of Site</i>	149
6.4.3 <i>Length of Stay, Social Behaviour and Stereotypic behaviour.....</i>	149
6.4.4 <i>Learning Ability and Stereotypic behaviour.....</i>	151
6.4.5 <i>Behaviour, Optimism and Use of Enrichment (Expt. 4)</i>	154
6.4.6 <i>Activity Budget</i>	154
6.4.7 <i>Stereotypic Behaviour</i>	155
6.4.8 <i>Optimism.....</i>	156
6.4.9 <i>Use of Enrichment.....</i>	158
6.4.10 <i>Aquisition of DogBox Task</i>	159
6.5 Discussion	161
6.5.1 <i>Effect of Site (Expt. 3)</i>	162
6.5.2 <i>Length of Stay, Learning Ability, Social Behaviour and Stereotypic behaviour (Expt. 3).....</i>	162
6.5.3 <i>Activity Budget (Expt. 4).....</i>	167
6.5.4 <i>Stereotypic Behaviour (Expt. 4).....</i>	168

6.5.5 Optimism (Expt. 4)	169
6.5.6 Use of Enrichment (Expt. 4)	172
6.5.7 Conclusions (Expt. 3)	174
6.5.8 Conclusions (Expt. 4)	174
7.0 Chapter Five	176
7.1 General Discussion	176
7.1.1 Potential Areas for Improvement.....	176
7.2 Main Findings.....	178
7.2.1 What is cognitive enrichment, and can it improve captive animal welfare?	178
7.2.2 Is cognitive enrichment more beneficial to captive animals than physical enrichment?	179
7.2.3 Is increasing environmental cognitive complexity associated with increasing welfare benefits?	181
7.2.4 Is enrichment better delivered in a group or individually accessed setting?	182
7.2.5 What affects the welfare of individually kennelled dogs? ...	183
7.2.6 Is it possible to develop an automated cognitive enrichment device that prevents habituation and is easy to implement in many laboratory, zoo and companion animal settings (e.g. kennel, home environment) that improves (and is also capable of measuring) welfare?	184
7.2.7 Can the welfare of kennelled dogs be improved by cognitive enrichment?	184
7.2.8 Is stereotypic behaviour a good measure of welfare and can it be reduced through cognitive enrichment?	185
7.2.9 Does the cognitive bias measurement paradigm suit this kind of research, what are the current weaknesses and how can they be improved upon?	186
7.3 Conclusions.....	189
7.4 Future Aims	190
8.0 References	192

1.1 Acknowledgements

Without the continued support of the following people, there is no way that I would have been able to indulge my love of animal welfare research to the extent that I have.

My two wonderful supervisors: Lisa Leaver who has been brave enough to leave me largely to my own devices as I worked through various crackpot schemes and ideas, and then give me exactly the right amount of guidance to leave me feeling confident in what I was doing and totally supported as a largely off-campus student, and Stephen Lea for being my mentor, advocate and tireless teacher, and for giving me the confidence as a fledgling masters student to carry on despite me sending a panicked email in the middle of the night deciding that I had made a huge mistake.

My brilliant parents, Nigel and Rosalyn, who I have leaned upon financially through my three decades of education (sorry!) while they supported my various geeky interests, thank you for putting in a few intensive weeks of babysitting while I shut myself in the study and did battle with my thesis. My lovely husband Roger who took it upon himself to marry a pregnant penniless PhD student at the point of writing up, and has been so supportive by believing in me and by taking our boy out for 'Daddy's Days' when I needed to concentrate. Our little boy Reuben for being the reason I am so motivated to get this thing written. My sister Sarah who has put up with hearing about my work more times than she probably cares to remember, and my in-laws Peter and Pauline for being there for so many days of baby-sitting duties for both Reuben and Roger as my housewife duties slipped and left everyone wondering what was for dinner again.

The RSPCA Little Valley Animal Shelter, Exeter and The Blue Cross, Tiverton, for being kind enough to allow me access to their dogs. My long-suffering assistant and friend Jo Gilbert, for helping me to collect hours and hours of dog data, highlights including sitting in a freezing cold kennel block with unmentionable smells and incessant noise, without complaining and only an occasional biscuit as payment. Catriona Ryan for instilling in me an interest in operant training and supporting my research. Finally, my 'animal' friends Lucy Halliday, Beth Nicholls, Emma Foster, Jess Isden, Kristin Thompson and Lucy Waite for being wonderfully weird and understanding me.

1.2 Declarations

- (a) I have followed the guidelines for the journal Applied Animal Behaviour Science for the four experiments carried out as part of this research and they are included in the form required for submission to the journal. The literature review will also be submitted. The journal was chosen for its high quality output in applied animal welfare and environmental enrichment for captive animals.
- (b) I have collated the reference sections for each part of the thesis into one section at the end for ease of reading.
- (c) A CD of scanned copies of data sheets, excel files of data, subject videos and other relevant materials is available on request. Due to a technical problem (deletion of hard drive from University server by IT department) a small amount of raw data from experiment one and two has been lost.
- (d) All work carried out under this research programme was considered a non-licensed project, which adhered to the Association for the Study of Animal Behaviour's Guidelines for the Use of Animals in Research and was fully disclosed to and permitted by the University of Exeter Ethical Review Group. It was covered by a full risk assessment for research carried out at the University of Exeter, the RSPCA (Exeter, UK) and the Blue Cross (Tiverton, UK).
- (e) The contributions of my supervisors to this thesis included a weekly meeting to discuss aims, plans, experimental design, data collection and analysis, as well as reading and commenting upon drafts. I conducted all of the research for the thesis myself except for experiment three in which the activity of the dogs was partly recorded by my assistant Jo Gilbert, who also carried out some cognitive bias task training with the dogs. I conducted all data analysis and wrote the entire thesis, and all views and opinions not otherwise attributed are my own. The literature review and two main data chapters will be submitted for publication, and will have Dr. Lisa Leaver and Prof. Stephen Lea as co-authors.

1.3 Ethics

In preference to restricting the rations of the pigeons in experiment one and two, I maintained all pigeon subjects at their optimum weight which varies from bird to bird – that is when a layer of fat can be felt over the breast bone and the pigeon is solid and robust, rather than their free-feeding weight which can be higher. They received their daily rations at the usual time of day, which was after the experiments were conducted. The dog subjects were also maintained at their optimum weight determined by a veterinarian at the shelters at which they were being cared for, and were fed before and after research took place each day. As this seemed to have little effect upon performance or willingness to participate by subjects, I suggest that maintaining at optimum weight may be an alternative to more restrictive diets if at all possible as welfare is likely to be improved (Lea, 1979).

No animal was forced to participate in any part of my research. Experiment one was carried out in the pigeons' home environment and experiment three and four were conducted in the dogs' kennels, but all subjects still received their full daily food ration if they did not choose to participate. In experiment two where group-living pigeons were given enrichment in an individual cage, the pigeons were released from their group housing and allowed the choice of whether to enter a holding cage (taken as volunteering) or fly back to the home cage. The pigeons did tend to choose to participate, probably due to the fact that a small reward was placed in each holding cage, so in effect they were actually 'paid' to participate. Any animal that became agitated during the experiment was removed from the research programme: three pigeons were removed from the cognitive bias task in experiment one and two because they did not attempt to learn the task and became withdrawn, and one dog seemed anxious during the observation part of experiment three so was removed from the study.

All work carried out under this research programme adhered to the Association for the Study of Animal Behaviour's Guidelines for the Use of Animals in Research, was fully disclosed to and permitted by the University of Exeter Ethical Review Group and was considered a non-licensed project (involving no invasive procedures). It was also covered by a full risk assessment

including research carried out at the University of Exeter, the RSPCA (Exeter, UK) and the Blue Cross (Tiverton, UK).

2.0 Chapter One

2.1 Literature review and general introduction: Does cognitive enrichment improve captive animal welfare?

Louise N. Millar*

**Corresponding author: Louise N. Millar, University of Exeter, School of Psychology, Washington Singer Laboratories, Streatham Campus, Perry Road, EX4 4QG. Email: L.N.Millar@exeter.ac.uk.*

Abstract

Animal welfare can be thought of as a continuum concept where the aim is not just the absence of negative affect, but the presence of positive affect. Welfare can be measured in many different ways, including activity budget assessments, physical changes, stereotypic and abnormal behaviours and cognitive bias, a relatively new paradigm that links operant response to ambiguous stimuli with animal emotion (Paul *et al.*, 2005; Bateson and Matheson, 2007). While taking a number of welfare measures helps to validate findings, the many combinations possible make comparisons between studies difficult.

The assessment of the value of environmental enrichment (EE), particularly cognitive enrichment (CE), for captive animals is hampered by the lack of consistent and methodical research. Welfare for captive animals in laboratory, farm, companion and zoo settings is difficult to assess and subject to interpretation. The perceived needs of animals change over time and according to fashion and policy. Quantitative, objective research on the effects of EE and other welfare improvement strategies is often partially achieved but lacking in long-term assessment or implementation. Different types of EE are thought to affect the welfare of captive animals in different ways depending on implementation strategy, species and other variables, and research into the effects of CE is particularly promising, with positive effects being seen in many species and settings (Manteuffel *et al.*, 2009; Clark, 2011; Franks, 2012).

The main conclusion to be drawn from this literature review is that CE is taking an increasingly central role in EE research, with evidence to suggest it is very effective at increasing positive behaviours and decreasing abnormal and stereotypic behaviours in farm and zoo animals. These findings must now be validated by methodical, objective research, including the cognitive bias welfare measure, as animal emotion, if we can properly access this, is an important part of the assessment of welfare: an animal's subjective experience of their environment is arguably more important than any other factor.

Keywords: Environmental enrichment, cognitive enrichment, animal welfare, captivity, stereotypic behaviour, cognitive bias.

2.2 Introduction

This review will cover the current state of animal welfare and environmental enrichment (EE) for captive animals, particularly cognitive enrichment (CE), and the welfare measurement techniques used to objectively evaluate environmental enrichment, particularly activity budget, physical condition, stereotypic and abnormal behaviour and operant techniques including cognitive bias measurement. Welfare and EE for captive animals in laboratory, farm, companion and zoo settings is hampered by the lack of consistent, methodical and comparable research. Particularly, there are shortcomings in current procedures for both the presentation of enrichment items and for measuring the welfare effects of provisions upon captive animals. After investigating these issues, the focus of my research will be introduced, specifically whether cognitive enrichment can be considered beneficial in terms of measurable welfare benefits for captive animals. Lastly, the questions I hope to answer experimentally will be put forward.

2.2.1 The Welfare of Captive Animals

Developing a Concept of Animal Welfare and Animal Rights

Descartes (1641; as translated by John Veitch, 1901) published in his document 'Meditations on First Philosophy' that non-human animals (hereafter animals) were like machines, they felt no pain, and had no ability to think. He proposed that they reacted unconsciously to stimuli. At that time, captive animals experienced isolated, cramped conditions, with no form of comfort or enrichment. They were operated on experimentally with no pain relief by educated men. Quite simply it was thought that the sole purpose of animals was to serve humans.

Some individuals such as Jean-Jacques Rousseau (1754) argued that animals should be afforded basic rights because they are sentient. In 1874, the Royal Society for the Prevention of Cruelty to Animals was founded, and minimum standards were introduced in farming and laboratories.

How Can Animal Welfare be Defined?

Broadly, a good state of animal welfare is evident if an individual animal is physically healthy, is not distressed, and behaves in a manner consistent with

innate drives. The World Organisation for Animal Health defines animal welfare in the document 'Terrestrial Animal Health Code' (2013) as 'how an animal is coping with the conditions in which it lives' (Chapter 7.1.1).

Poor welfare can be defined as a negative affective state, found in individuals experiencing chronic stress for various reasons, and is associated in both humans and animals with increased mortality through continual activation of the physical and psychological stress systems, specifically the hypothalamic-pituitary-adrenal and sympathetic-adreno-medullary axes (Walker *et al.* 2012). In other words, the experience of long-term stress indicated by the individual's affective response (such as anxiety or depression) to possible stressors (e.g. in animals, physical discomfort due to ill-health, insufficient food, social isolation, barren cramped captive environment, unpredictable, aversive husbandry) is directly linked with a shorter lifespan in animals.

Animal welfare can be thought of as a continuum concept where the aim is not just the absence of negative affect, but the presence of positive affect. (Duncan, 1996; Boissy *et al.* (2007b). The possibility that animals experience emotions as we do, if at all, is still controversial. However, Boissy *et al.* (2007b), Panksepp (2011), Panksepp and Lahvis (2011) and Leliveld *et al.* (2013) suggest that at least some animals share the physiological, chemical and behavioural markers of emotion with humans and therefore it is likely that they also share the emotional feeling even if not the conscious cognitive component. Recent advances in the study of animal emotion (e.g. Bateson and Matheson, 2007; Mendl *et al.* 2009) have found that affect can influence cognitive processes in animals.

It should be assumed that an individual animal can experience welfare subjectively (Dawkins, 1990), and differently to other animals experiencing the same physical conditions, requiring that we assess welfare continuously and on an individual level.

Responsibilities of Care-Givers

According to the UK Animal Welfare Act (2006), any person responsible for captive animals has a duty to make sure all their needs are met, and if they fail to provide this level of care, they are deemed to have committed an offence. These needs as summarised from the Act are as follows: an animal must be protected from pain, injury, suffering and disease, be provided with suitable food

and water, be given contact with other animals (if appropriate), be housed in a suitable environment, and be able to exhibit normal behaviour patterns. These last two requirements are extremely vague, and subject to various interpretations which can shift over time.

This puts us, as wardens of captive animals, in a position of responsibility to ensure their experience of life is as positive as possible (Boissy *et al.* 2007b), and beyond the basic provisions of food, water, shelter and good physical health, we can ensure that the captive environment is also optimised for animal welfare through environmental enrichment.

2.2.2 Environmental Enrichment

What is Environmental Enrichment?

Improvements to the environment of captive animals which go beyond the most basic provisions of food and water are known as environmental enrichment (hereafter EE). The requirements that animals be housed in a suitable environment, and be able to exhibit normal behaviour patterns, are very difficult for animal keepers to comply with in a captive setting. Wemelsfelder (1997), Bloomsmith *et al.* (2007) and Shuangying *et al.* (2009) argue that the captive environment can be considered damaging by its very nature, which is typically unstimulating, unchanging and unchallenging. Natural behaviours, including flying, running, climbing, digging, nesting, bathing, hiding, foraging for a varied diet, and interacting with other animals, require housing modifications that allow such behaviours to occur. For this reason, research is conducted with zoo, laboratory, companion and farm animals in order to pinpoint which kinds of EE are actually valuable to animals, and which are simply aesthetically pleasing to their human keepers. Newberry (1995) suggests that enrichment that facilitates adaptation to the specific environment and situation that an animal experiences is more important than retaining 'naturalness'. Franks (2012) argues that animal welfare is a dynamic process rather than an animal's passive acceptance of the environment and situation in which they find themselves.

How Does Environmental Enrichment Affect Animal Welfare?

Fox *et al.* (2006) propose that effective EE has important consequences for the neurobiology, physiology and behaviour of captive animals. They

reviewed over 100 studies, mainly conducted upon laboratory mice and rats, that revealed measurable physical and neurobiological effects of EE on the animals, as well as significant behavioural changes. These studies found that the benefits of EE include reduced anxiety, increased learning abilities, better body condition, more resistant immune systems, faster recovery from illness, and less fearfulness and hyperreactivity in new situations. Fox *et al.* (2006) argue that the isolation of social animals has a particularly detrimental effect, and should be considered when planning animal housing. They suggest that minor stressors such as daily handling, puzzle-solving or training can be positive for animal welfare, especially if they are predictable. Animals that are given EE tend to be more emotionally stable and exploratory, behave less fearfully and have less adrenal response even when given electric shocks. Fox *et al.* (2006) argue that the reason for this is that the experience of EE can provide a physiological and psychological buffer against stressors which would otherwise have a detrimental effect on welfare, working like a vaccine – it is in fact referred to as ‘stress inoculation’.

Csatadi *et al.* (2008) looked at the effects of introducing enrichment to a group of captive bonobos which had developed a negative behaviour (stealing and rough handling of an infant). They noted that enrichment that worked at a group level e.g. a hanging feeder was particularly effective, while individual enrichment e.g. a food ball which only a single individual could dominate, was less effective. They argue that animals which are rare in the wild e.g. bonobos, should have extra efforts made with their captive care, as they have more resting on their successful living and breeding.

Fox *et al.* (2006) suggest that control seems to be a very important factor in the effectiveness of EE: if the animal is able to control aspects of its environment, they are likely to be able to cope better with challenges and welfare is likely to be increased. Latham and Mason (2010) agree that there are beneficial effects of EE, finding that animals raised in an enriched environment, which enables them to behave in ways they are naturally motivated to, display far less stereotypic behaviour, and have lower corticosteroid levels than animals raised in barren environments, suggesting that enriched animals are less stressed by captivity. The benefits seem to be highest when animals are provided with EE from birth, but EE introduced later also seems to have measurable benefits.

EE and Intrinsic Behavioural Needs

Animals have many intrinsic needs such as finding shelter, foraging, attracting and keeping a mate, and producing surviving offspring. Captive animals that cannot satiate these intrinsic needs due to deficits in their environment often suffer from poor welfare. Here foraging is used as an example of a behavioural need that is often frustrated by captivity.

Foraging is not merely the act of consuming food, it is a complex behaviour involving searching, acquiring and processing food items of different values and involving different risks, and as such takes up much of a wild animal's daily activity budget. In contrast, captive animals often have their food presented to them in a bowl, reducing their foraging time to a few minutes per day. Animals seem naturally motivated to explore and forage in their environment, even when food and other resources are freely available, a concept known as contrafreeloading first systematically demonstrated by Neuringer (1969) and hotly debated by many researchers since (e.g. Osbourne, 1977; Inglis *et al.*, 1997). As Inglis *et al.* (1997) propose, contrafreeloading behaviour maintains the ability to obtain food using that technique if it subsequently becomes the only source. Wemelsfelder (1997) adds the notion of 'agency', of being an active agent in one's own survival, and this means that animals have an intrinsic drive to explore, interact with and control the environment, which has an evolutionary benefit. Recently, in a more naturalistic experiment than the traditional operant method of investigating contrafreeloading, Vasconcellos *et al.* (2012) found that captive maned wolves preferred to find food scattered in a forested area in the presence of freely available food from a tray, supporting the notion that there is an intrinsic need for animals to use their natural foraging abilities even when they are not immediately needed for survival and to be in control of their environment. This example therefore suggests that by using EE to increase the time it takes for captive animals to forage for their daily intake of food, welfare may be improved.

Current Challenges in Environmental Enrichment Research and Implementation

Newberry (1995), Young (2003), Fox *et al.* (2006), Meehan and Mench (2007), Azevedo *et al.* (2007) and Adams (2007) all agree that there is still very little consensus when it comes to defining EE. Newberry (1995) defined EE as

any modification to an animal's environment which leads to an improvement to the biological functioning of the animal. This is a very broad definition. What constitutes EE is a difficult point: in some studies it is extra space (e.g. cows, Jensen *et al.*, 1998; Tapki *et al.*, 2006), physical structures (e.g. mice, Pietropaolo *et al.*, 2004), food (e.g. rats, Moncek *et al.*, 2004; dogs, Schipper *et al.*, 2008; maned wolves, Vasconcellos *et al.*, 2009), bedding (e.g. mice, Van de Weerd *et al.*, 1998), substrate (e.g. chickens, Shields *et al.*, 2004), or toys (e.g. pigs, Smith *et al.*, 2009), in others it is puzzle feeders (e.g. chimpanzees, Celli *et al.*, 2003), or social interaction (e.g. cows, Jensen *et al.*, 1998; foxes, Hovland *et al.*, 2011). In some cases, enrichment items are renewed or replaced with different items regularly, while in others they are not, and time spent with enriching items varies from a few minutes to continuous exposure (Fox *et al.*, 2006). The type of enrichment likely to be important to rats may have no effect on great apes or elephants. This means that each type of EE has to be evaluated in a methodical way at a species level as well as at group level and even at an individual level before it can be said to enhance welfare.

Newberry (1995) complains about the use of the word 'enrichment' being used in studies as a description of what has been done by experimenters, e.g. increasing the complexity of the environment, leading to a false positive impression of the results. Enrichment, argues Newberry (1995), should be used to describe the *effect* upon an animal's experience of a captive environment, i.e. is the animal enriched by the provisions made? Perhaps the phrase 'environmental enrichment' should only refer to situations when an animal's environment has been enriched in some way that actually causes a measureable welfare improvement; otherwise it is simply furniture or an aesthetic improvement for human observers, an environmental elaboration.

Standardization of EE is clearly needed to facilitate meaningful comparisons, and then a systematic evaluation of each type of EE should be carried out to gain a better understanding of their relative value. This is a massive task and has not yet been undertaken to any significant degree. Wells (2009) suggests that any enrichment needs to be carefully assessed in terms of the costs and benefits to the animals, and preferences at species, group and individual levels should be taken into account. A thorough investigation of the behavioural, physiological and affective changes associated with the introduction of any item is needed.

2.2.3 Types of Environmental Enrichment

Bloomsmith *et al.* (1991) identified five main areas of environmental enrichment (EE): Social (for example interaction with conspecifics), Occupational (for example puzzles, running wheels), Physical (for example size and complexity of enclosure, accessories such as ropes), Sensory (for example sounds, smells, visual stimulation) and Nutritional (for example types of food available and presentation of food in addition to their basic ration, such as live insects, fresh fruit, raw food, nuts in shells, frozen berries, small items such as seeds etc.). In my view some reclassification is required since the enrichment value particularly of puzzles and running wheels is based on very different motivations. I suggest the reclassification to be as follows: Social (as before); Sensory (as before); Nutritional (as before - food can also be the primary reinforcer for a wide range of enrichment so the *design* of enrichment rather than the presence of food should be the classifying factor); Physical, focussing upon providing animals with physical tasks, for example scattering food around an enclosure so that it takes longer to collect, or providing items such as branches, ropes, substrates, running wheels, bedding material and shelters in order to encourage a range of natural behaviours; and Cognitive, focussing on enriching captive animals by challenging their cognitive abilities such as learning and memory through puzzles and positive reinforcement training. What follows is a brief review of social, sensory, physical and nutritional enrichment and an in-depth review of cognitive enrichment, although it must be noted that each area represents a vast amount of research and as such cannot be reduced sufficiently – in each case suitable review works are indicated if available for further reading. For an overview, Honess and Marin (2006) provide a good review of the different types of EE available and the effects they have on primate welfare particularly.

Social Enrichment – A Brief Review

Rault (2012) reviewed the positive effects that social support from conspecifics has on farm animals exposed to stressors and found evidence from over 150 studies to suggest that contact with familiar conspecifics is beneficial for the welfare of farm animals in terms of better physical (e.g. increased immunity and better recovery from illness) and psychological response (e.g. reduced anxiety, better recovery from agonistic encounters) to

stressors. Far from the risk of agonistic behaviour being a reason for individual housing, individuals in a stable social group can cope far better with stress, uncertainty and other situations likely to reduce welfare than isolated individuals. Beerda *et al.* (1999) conducted a study with dogs that were housed in social groups then moved to individual kennels. They found that the dogs housed individually behaved in ways consistent with a chronic stress response, showing increased autogrooming, excessive barking and stereotypic behaviour as well as increased anxiety and aggressive behaviour when subsequently faced with an unknown conspecific compared to baseline behaviour from when socially housed. This suggests that moving from a social environment to an isolated one is chronically detrimental to animal welfare. Hovland *et al.* (2011) measured the strength of female vixens' motivation to visit another female vixen and found that the cost they were willing to pay (lever pulls) was about a quarter of the cost they were willing to pay for food, and this varied with dominance suggesting that for subordinate individuals, social contact was a less attractive prospect than for dominant ones, especially if there was no opportunity to escape. Honess and Marin (2006) discussed the relative benefits of social housing for non-human primates (hereafter primates) in their review of enrichment techniques for primates and also found in favour of social housing in zoo and laboratory settings at least in pairs although again dominance and sexual competition must be considered when making grouping decisions. In addition, Honess and Marin (2006) note that abnormal behaviours in primates raised in isolated conditions are difficult to eradicate but social housing at least reduces its prevalence. The above research suggests that for any housing system there should be evaluation on the basis of each individual animal's experience, but that social housing should be the basic protocol for all captive animals. Rather than being an enrichment in the sense of an improvement upon the basic provisions for survival, social contact with conspecifics should be considered a fundamental right for captive animals.

Social contact with humans has also been found to be beneficial to at least some captive animals: Shiverdecker *et al.* (2013) found that shelter dogs that were given thirty minutes of social contact with humans within forty hours of first entering the shelter exhibited significantly lower blood cortisol levels than control dogs that did not experience this human contact, suggesting that the dogs found it reassuring in a new environment. It is unlikely that wild-caught

animals of non-domesticated species would benefit from human contact as much as domestic animals would, so careful assessment on an individual basis is needed.

Sensory Enrichment – A Brief Review

Wells (2009) provides a good review of sensory enrichment for captive animals. A study by Shuangying *et al.* (2009) found that even a simple enrichment such as introducing prey odours into a predator animal's environment was enough to increase behavioural diversity significantly, although only for a relatively short time. Graham *et al.* (2005a) placed different essential oil fragrances into the kennels of shelter dogs and found there were significant behavioural effects consistent with relaxation including increased resting and reduced barking, particularly with lavender and chamomile scents. However, Wells *et al.* (2007) investigated the effect that essential oil scents had upon zoo-housed gorillas and found that they had little impact in terms of behavioural change.

Sherwin and Glen (2003) found that mice housed in different colour cages showed a definite preference for the colour white and dislike of red – mice housed in red cages also showed significantly more behaviours consistent with anxiety suggesting that colours in the environment may have an effect on animal welfare.

Wells *et al.* (2002b) found that shelter dogs that were exposed to classical music in their kennels showed behaviour consistent with increased relaxation including reduced barking and increased resting. Wells *et al.* (2006) found that zoo-housed gorillas that experienced classical music showed large reductions in abnormal and aggressive behaviour compared with baseline measures. The effect was larger than when the gorillas were exposed to ecologically relevant sounds (rainforest noises). Wells (2009) suggested that sensory enrichment such as music may be beneficial as a mask from outside noise or distraction from boredom, but animals may only really benefit if they can choose to turn it on and off, and it is important to note that 'natural' sounds, i.e. rainforest noises, can in fact scare animals who have never lived in the wild (Wells, 2009).

Wells (2009) also discussed the effects of moving visual images such as videos on welfare, the consensus being that positive behavioural changes such

as reduced stereotypic behaviour and fear response have been observed in a wide range of species due to exposure to video images. Moving images of biologically relevant stimuli such as prey or conspecifics have been well-received - Bloomsmith *et al.* (1990b) found that chimpanzees preferentially watched videos of conspecifics when given the opportunity, individually housed chimpanzees showed no sign of habituation to video images, and undesirable behaviours reduced. However, Schapiro and Bloomsmith (1995) found that rhesus macaques were not interested in watching videotapes of conspecifics, and when Honess and Marin (2006) reviewed the literature regarding visual stimulation such as video images for primates, they concluded that while primates do seem to look at video images, there is not much evidence that it actually improves welfare. Wells (2009) argues that it is the biological relevance, not the naturalness of the enrichment that matters – domestic cats have been found to enjoy watching a video of a snooker game as the fast, linear movements of the balls mimic their natural prey (Ellis and Wells, 2008).

These findings suggest that sensory enrichment has some value as enrichment for captive animals and the effects of different types should be carefully evaluated as they are based on different physical systems, with different species valuing the input from different senses varyingly.

Physical Enrichment – A Brief Review

Physical enrichment devices such as toys, with no food reinforcer, are typically used by captive animals for less than a day before habituation (Young, 2003). Newberry (1995) argues that the use of the word 'toys' suggests that adult animals are motivated to play with objects, and while this is occasionally observed, animals generally gain greater benefit from objects if they provide a specific benefit to the animal, e.g. the opportunity to climb, hide, wear down fast-growing teeth and claws, or obtain food, and that in addition, animals that approach and investigate an object are not necessarily experiencing increased welfare due to its presence. I would introduce the proviso that perhaps this works the other way around also, and interaction is not essential for EE to improve welfare, for example a view from a window does not need to be interacted with in order for the feeling of space and light to improve one's mood.

Foraging and grooming boards, where turf, fleece or other materials is attached to a board and animals search through to find food or other items, are

typically well-used by primates, and are thought to work on the principle of providing an outlet for natural foraging urges and increasing search time for food. However they are not always successful in improving welfare as in some cases aggression, stereotypic behaviour and abnormal behaviours are also seen when these devices are provided (Honest and Marin, 2006).

Nutritional Enrichment – A Brief Review

Nutritional enrichment relates to the provision of food in an interesting way, requiring increased processing time, such as nuts in shells, frozen food, small food particles and species-appropriate 'treat food' such as cheese, insects, berries or honey. Beirise and Reinhardt (1992) found that providing monkeys with food that required processing such as nuts in shells and unhusked corn increased total foraging and processing time and increases in species-typical behaviour. Schipper *et al.* (2008) found that kennelled dogs given Kong™ feeding enrichment toys that promoted extended foraging and chewing behaviour, barked less and exercised more than dogs in a control group fed only normal commercial dog food. This suggests that food that is interesting in texture and flavour and presented in a way that takes longer to obtain than regular food can be a useful, easy to implement EE and has measureable effects on behaviour and welfare. Honest and Marin (2006) warn against predictable scheduling of nutritional enrichment as stereotypic behaviour and other abnormal behaviours can arise.

2.2.4 Cognitive Enrichment – An In-Depth Review

As early as 1925, Yerkes postulated that the most promising avenue for enrichment at least with captive primates is the design and implementation of apparatus which can be used for play or work (Yerkes, 1925). Kiley-Worthington (1983) suggested that captive animals should be given the opportunity to work for rewards in order to overcome boredom. Cognitive enrichment (CE) for animals is a term originally coined by Milgram (2003) when describing a battery of cognitive tasks given to dogs, found to slow age-dependent cognitive decline. Sale *et al.* (2009) recently reviewed the area of literature concerning EE and brain function, and found that engaging with complex cognitive tasks has the capacity to modify behaviour, enhance learning and memory, and reduce ageing-related cognitive decline in animals. Manteuffel *et al.* (2009) define CE

as any type of enrichment that requires cognitive effort from the animal, leading to an ability to exert active control of the environment. Hall *et al.* (2009) found that in humans, regularly doing crossword puzzles delays the onset of cognitive decline in dementia. The reason they give for this is that cognitive activities contribute to cognitive reserve, or the level of resistance of the brain to neurological damage. Lustig *et al.* (2009) reviewed the neuroimaging data available on the effects of brain-training and found several areas of the brain that are responsible for core cognitive processes that may be 'trained' in transferable skills. There is now a range of electronic games on the market aimed at 'brain-training' or improving your cognitive abilities. Their theory is that the brain is like a muscle that needs exercise or it will decrease in power, reducing the ability of the brain to cope with demands such as problem-solving, long and short term memory retrieval and learning new things. However, Owen *et al.* (2010) conducted a large empirical study of these claims and they suggest that while training on cognitive tasks improves skills in the specific tasks being trained as expected, these benefits are not linked with a general improvement in cognitive function.

The Yerkes-Dodson law of arousal and performance states that as levels of arousal increase, so does performance on a task, up to a point at which performance begins to decrease. This can be represented graphically as an inverted U-shaped curve showing arousal along the x-axis and performance on a task along the y-axis (Yerkes and Dodson, 1908). Redding *et al.* (1988) found in their study of children given cognitive puzzles that a moderate challenge seemed to be more pleasing, and elicit more persistence, than either an easy or difficult puzzle, suggesting that even two-year old children enjoy mastering a task, and are intrinsically motivated to do so. This motivation is of evolutionary benefit, as a child interested in mastering challenges in the environment is likely to succeed and survive as long as the challenges are not insurmountable. Franks (2012) likens the concept of appropriate challenge in animal welfare to that of children engaging in risky behaviours that allow them to develop skills and self-confidence to cope with other situations – to become 'effective'.

The concept of 'Flow' is explained by Csikszentmihalyi (1990), and describes the state of positive, satisfied emotion when a difficult challenge can be met with appropriate skill. Fig. 1 demonstrates this phenomenon, which has been used to describe human experience but can also be applied to animal

welfare (Meehan and Mench, 2007; Clark, 2011). Laule and Whitaker (2001) argue that apes ‘appreciate the opportunity to learn something and to achieve competence’ when reviewing studies of ape EE.

A mismatch of challenge and skill can lead to anxiety (if the challenge is too much for the skill level of the animal), apathy (when neither the challenge nor the skills of the animal are up to much) or boredom (when the animal’s skills outrank the available challenges). Leavens *et al.* (2001) found that chimpanzees displayed more self-directed behaviours such as scratching when exposed to more difficult cognitive tasks than to easier tasks, suggesting they were mildly stressed by the challenge.

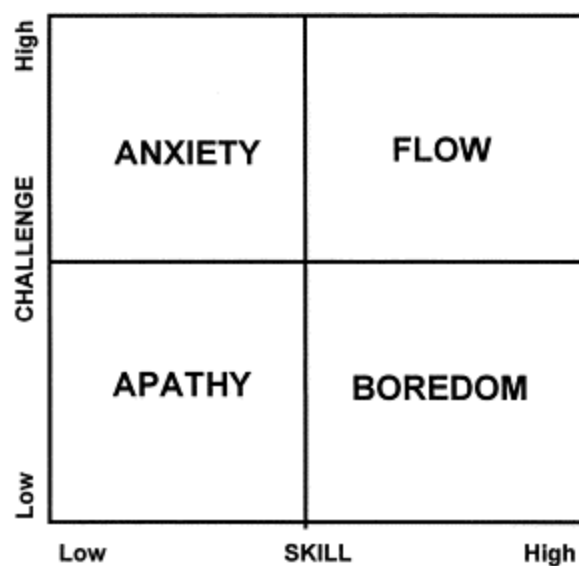


Fig. 1: The outcomes of varying levels of challenge and skills. (Diagram reproduced from Chen and Wigand, 1999).

Myers and Diener (1995) evaluated several life elements as intrinsic to humans achieving happiness. A happy person, according to Myers and Diener (1995), has a sense of control over their own lives, is optimistic, has a social support network consisting of significant others such as spouse, family and friends, and engages in meaningful, challenging work. Causality is almost impossible to untangle: is a person happier because they are optimistic and in control, have a strong social network and engage in meaningful work, or is it the other way around? A person with a happier affect would perhaps seem friendlier and more competent, be given better opportunities in life and therefore get a

better job. The assessment of 'happiness' in animals is even more difficult, but at the very least, the correlates of happiness found by Myers and Diener (1995) suggest that benefits may be felt by animals too if they are provided with opportunities to control their environment, to engage with other conspecifics if possible, and challenging tasks to do.

When designing and testing EE it is important to factor in the individual cognitive abilities of the animal. In order to survive in the wild, animals must use their cognitive skills, such as navigation, memory, perception, learning, recognition, categorization, problem-solving, decision-making, reasoning, as well as sometimes very complex social cognition such as hierarchical behavior, reconciliation and pack-hunting. Pepperberg (2004) highlighted this need for cognitive challenges to replace wild experiences in order to prevent boredom and unsuitable behaviours developing, based on her work with grey parrots (*Psittacus erithacus*), and extrapolated it to other captive zoo and companion animals. Wechsler and Lea (2007) also suggest investigating the possibility of providing agricultural animals with cognitive tasks to reduce boredom in captive situations.

However, Meehan and Mench (2007) and Manteuffel *et al.* (2009) claim that the current trend for EE is to satisfy only the physical needs of an animal, for example through inanimate appetitive or structural applications, which have the disadvantage of being very quickly habituated to, or by scatter-feeding to increase the time spent foraging, known as physical enrichment (PE) and suggest that there is a need for research into the effects of appropriate challenge on the welfare of captive animals. Azevedo *et al.* (2007) conducted an analysis of the studies conducted on environmental enrichment for captive animals up to the time of their investigation, and found that nutritional, sensory, social and physical enrichments were all well-represented across most of the animal taxa, while CE was used in less than 4% of mammal studies, and in no bird, fish, invertebrate, reptile or amphibian study assessed by the analysis. There are now several researchers seeking to rectify this deficit. The Research Institute for Farm Animal Research based in Dummerdorf, Germany has been studying the effect of CE on farm animal welfare. Great ape welfare is currently being examined in terms of the benefits of appropriate cognitive challenge by Clark (2011), who puts forward the argument that whilst cognitive challenge should not be seen as a 'cure-all' for apes in captivity experiencing poor

welfare, its effects should certainly be investigated as it may alleviate some of the effects of captivity on ape welfare. Meanwhile, Franks (2012) has begun looking into the benefits of cognitive challenge for both humans and captive rats, and has found that engaging in effectiveness while successfully coping with cognitive challenge improved both human and captive rat welfare, and suggests that according to her 'engaging in effectiveness' model, being exposed to challenges should increase an individual's general ability to meet future challenges, and that appropriate challenge is a vitally important enrichment consideration.

Markowitz and Line (1989) suggested that EE that allows both control by the animal and response to the animal may be less habituated to and used more than inanimate stimuli. Control has been found to be important for captive animal welfare by many studies, including Franks (2012) and Clark (2011). As Franks (2012) points out, it is the successful completion of a challenging task that is enriching, rather than just the availability of such tasks, and accordingly animals provided with CE should be cognitively, behaviourally and physically able to face those challenges. Langbein *et al.* (2009) suggest that giving captive animals the opportunity to use their cognitive abilities by presenting them with CE, and allowing them to control their environment to some extent, should be considered of great importance in increasing their psychological well-being.

Captive animals can begin to show undesirable behaviour when they are not challenged enough. Daily enrichment should be provided in order to prevent boredom and habituation, and well-designed enrichment should stimulate captive animals both mentally and physically, and hold their interest (Csatadi *et al.*, 2008). Clark (2011) suggests that the most important things to consider when designing cognitively challenging EE (CE) are an appropriate level of challenge according to the animal's skillset or learning capabilities, aiming to lead to feelings of control and 'flow' (possibly measurable by testing the level of 'distractability' from the task) for the animal, and an 'ejector seat button' or way to escape if the animal feels overwhelmed by the challenge. Wechsler and Lea (2007) argue that cognitive tasks given as EE should be implemented without the need for human trainers in order to reduce the cost and burden upon animal carers.

CE in Action

Manteuffel *et al.* (2009) describe in their review of cognitive enrichment techniques for farm animals how successfully coping with a cognitive challenge has been found in pigs to be intrinsically rewarded by increased dopamine production, and suggest that stereotypic behaviour may be reduced by the introduction of CE, due to the animals being less bored and frustrated by their environment. They suggest that the long-term integration of operant devices into the home environment of captive animals seems to be more beneficial in terms of welfare as opposed to shorter-term enrichment strategies.

An ongoing research programme at Disney's 'The Seas' facility has shown that dolphins, with their highly developed cognitive skills, complex social behaviour, echolocation abilities and foraging techniques, have a need to engage in cognitive tasks in captivity in order to achieve good welfare. The group of dolphins had been used intensively since 1988 to further knowledge about dolphin cognitive abilities and behaviour which would be difficult to collect in the wild. Harley *et al.* (2010) reviewed the cognitive research programme being undertaken at the facility and concluded that the dolphins benefit from the research and will participate both when unwell and when no food rewards are given, suggesting a high level of motivation to engage in cognitive challenge.

Tarou and Bashaw (2007) stated that extrinsic reinforcement is important in preventing habituation – they refer to the study carried out by Platt and Novak (1997) where rhesus monkeys were presented with both a food-reinforced video game and a video by itself. The monkeys habituated to the video images eventually, but less quickly to the extrinsically reinforced game. However, it could be argued that the video game was more intrinsically rewarding than the video images because it was interactive. Platt and Novak (1997) argue that in their study, reinforced video games were found to be enriching to the rhesus macaque subjects, and therefore can be used concurrently as research tools and enrichment. However, Tarou and Bashaw (2007) argue against continuous reinforcement, and suggest instead that using a partial reinforcement schedule is a very effective method of eliciting long-term use of enrichment, in terms of reducing habituation, satiation and possible extinction if the rewards run out before the session is over.

Langbein *et al.* (2009) conducted a study with group-housed goats, examining the possibility that goats seek and benefit from cognitive challenge.

Correct responses to stimuli in the operant device were rewarded by sips of water. They found that the goats continued to engage with the task even when water was available freely elsewhere and suggest that this is evidence that the goats found cognitive challenge intrinsically rewarding.

Hagen and Broom (2004) found that cows given a simple cognitive task showed emotional reactions to their own learning, behaving more excitedly with increased jumping, bucking and kicking when they improved upon a task compared to a yoked control group fed on the same temporal schedule. It is suggested by Hagen and Broom (2004) that animals engaging in operant training can show excitement when they anticipate a reward, and also when they have some control over the appearance of the reward. This has implications for cognitive tasks being good for welfare but only if the animal feels in control of the outcome.

CE Considerations

According to the findings reviewed above, therefore, a CE paradigm should involve an extrinsically reinforced interactive device, which the subject can control, with a partial reinforcement schedule, presented on a rotation with other enrichment, in an animal's home environment, which provides an optimum level of cognitive challenge.

Meehan and Mench (2007) and Clark (2011) argue that designing cognitive challenges that fit the above criteria is no easy task and should not be undertaken lightly, as a mismatch may lead to lower welfare than the animal started with. Yamanashi and Matsuzawa (2010) found that far from there being a 'one-size-fits-all' approach to cognitive task difficulty even within a species, the level of anxiety displayed by six chimpanzees (measured using self-directed-behaviours) during cognitive tasks varied significantly depending on the individual, which Yamanashi and Matsuzawa categorized as sensitive or non-sensitive. This categorization seemed to be unrelated to cognitive ability, and also did not follow a family lineage in the study. Therefore, the stress levels of animals undertaking cognitive challenges need to be monitored, otherwise their welfare may suffer from the enrichments intended to improve it. In addition, if these animals are being used in cognitive experimentation, their ability to cope with challenge may affect their usefulness as a test subject.

Platt and Nowak (1997) and Langbein *et al.* (2004) argued that it is important that cognitive tasks intended to improve welfare should be accessible from an animal's home environment, in order to keep stressful movement and interventions to a minimum. In a group-housed setting, the problem of separating animals in order to provide cognitive enrichment in isolation has been discussed by Manteuffel *et al.* (2009), who argue that it is not beneficial to isolate group-housed animals who may become stressed by the separation. Instead, they suggest that the provision of several automated learning devices, personalised training criteria (e.g. auditory command for an individual to attend the device), or access to the device restricted to one animal at a time, would be less stressful. A bonus of providing CE in an animal's home environment is that while traditional cognitive testing is carried out in a laboratory and animals are usually put in to operant devices for the purpose, a more relaxed animal accessing the equipment voluntarily is likely to give a better indication of the cognitive abilities of that animal. Fagot and Paleressompouille (2009) gave baboons voluntary access to cognitive testing equipment twenty-four hours a day in their home environment. The baboons showed high levels of cognitive achievement by learning complicated cognitive puzzles, and indicated their interest by their high level of voluntary participation.

In this context, the provision of CE is particularly important for laboratory animals used for cognitive experimentation: if animals kept in a state of cognitive deprivation are used in cognitive experiments where the key is getting animals to learn, it is likely that those animals will provide lower-grade results than animals kept in cognitively stimulating environments. Newberry (1995) was concerned with the use of animals in cognitive studies that are kept in suboptimal conditions, e.g. the standard laboratory cage – specifically the validity and applicability of the findings of cognitive studies using animals that do not have access to cognitive challenge in their home environment. Poole (1997) reviewed the evidence regarding the effect that the psychological state of animals has on their behaviour, physiology, immunology and other parameters, concluding that a happy (psychologically as undamaged by captivity as possible) animal is the optimum research model and this should be aimed for by scientists using animals for any purpose in order to conclude reliable findings from their research. Fox *et al.* (2006) found that studies using rats in cognitively challenging maze experiments showed the animals performed

better and were less stressed when they had been kept in EE housing. Hausberger *et al.* (2007) found that horses that performed stereotypic behaviour due to suboptimal environmental conditions were less able to learn an operant task and if they did, took longer to learn it, than horses with no stereotypic tendencies. The use of laboratory animals in cognitive experimentation is based on the assumptions that (a) the animals are fit to work under the conditions they are kept, and (b) they will provide reliable, valid results. Neither assumption is justified if animals are kept under unsuitable conditions. In addition, if cognitive enrichment is beneficial to laboratory animals used in cognitive research, there is the possibility that far from lowering the welfare of the animals, the experiments themselves are increasing the welfare of the animals used. This is an important consideration in the use of animals for cognitive research. As Platt and Nowak (1997) asserted, researchers may be able to gather cognitive data from subjects while simultaneously improving their welfare.

Types of CE

Puzzle Feeders

A puzzle feeder is usually a device containing food items that provides a cognitive challenge as it must be manipulated in some way by learning to operate doors, levers, buttons etc. in a certain way in order to release a food reward. A meta-review of zoo mammal enrichment and stereotypic behaviour by Shyne (2006) found that although there are many problems inherent in zoo research including small subject numbers, the nature of subjective reports of enrichment effects and lack of statistical knowledge among zoo staff, there was a clear trend of stereotypic behaviour being reduced more by the introduction of 'food puzzles' than by other types of enrichment such as structural changes or furniture. While this is a vague description, it is likely that most of these devices required a cognitive puzzle to be solved in order for subjects to obtain food.

Celli *et al.* (2003) found that chimpanzees given puzzles requiring tool use to extract honey increased their foraging time significantly, which was taken as evidence of welfare improvement as their activity budget was more analogous to that of their wild counterparts. Dominance affected access to the devices and consequently dominant individuals benefitted more from the devices than subordinates. Honess and Marin (2006) reviewed approximately

thirty-six studies evaluating the use of puzzle-feeders with primates and concluded that their use increases foraging time, and reduces abnormal and stereotypic behaviours. This suggests that cognitive enrichment has a history of being more effective than other types of enrichment at improving welfare as indicated by reduced stereotypic behaviour. The effect of puzzle feeders on aggressive behaviour is less clear, with some studies reporting increases in aggression while others reported decreases (see Honess and Marin, 2006). If a puzzle feeder is large enough for all animals to use it at the same time, or there are enough devices for all animals, aggression may be reduced. In addition, if there is a dominance hierarchy that prevents subordinate animals from accessing the device, a property of being able to scrounge from successful animals (perhaps by making rewards abundant, small and easy to drop) may increase the welfare of subordinate animals as well as the more successful ones (Honess and Marin, 2006).

Positive Reinforcement Training (PRT) with human trainer

Coppola *et al.* (2006) found that PRT significantly reduced cortisol levels of dogs newly entering a shelter environment. Dogs were taken out of their kennel on the second day of their stay to interact with a human trainer for approximately forty-five minutes, during which time several basic training commands were practiced. The next day these dogs had lower cortisol levels than a control group of dogs that had not experienced the contact and training time with a human.

Bloomsmith *et al.* (2007) examined the use of PRT with primates and found that even severe stereotypic and abnormal behaviour can be moderated through the use of PRT aimed at reducing those behaviours and replacing them with more suitable, species-appropriate ones, making PRT with primates analogous to some psychiatric methods used in human behavioural therapy. Pomerantz and Terkel (2009) found that chimpanzees trained to present body parts on command using human-led PRT showed a decrease in abnormal behaviours and increases in affiliative behaviours unrelated to the behaviours being trained, lasting through the day. Subordinate individuals showed more welfare improvements than dominant individuals, which could have been caused by increasing their perceived control over the environment or simply by being given more attention and stimulation by keepers. Pomerantz and Terkel

(2009) conclude by describing PRT as an effective EE tool with demonstrable psychological benefits that can and should be used with captive animals.

Operant tasks

Langbein *et al.* (2004) used a visual discrimination task in an 'automated learning device' – a pen with a screen at one end and buttons for the subjects to press, more commonly known as an operant device. In the study, goats could access the device voluntarily as it was installed in the goats' home pen which meant that there was no additional stress from having to leave the home environment. They found that heart rate varied according to learning in the goats – a higher heart rate indicated a challenge and excitement from learning that a reward follows a correct choice, followed by a lower heart rate as the animals relaxed and achieved 'flow' in their learning activity. Langbein *et al.* (2004) argued that animals need a way to escape boredom and feel in control of their environment, and their study with goats showed that cognitive training allowed animals to benefit from learning a challenging task without it becoming a stressor itself.

Puppe *et al.* (2007) examined the effect of a novel cognitive task given to growing pigs, in this case learned acoustic cues to signal access to different feeding stations. They found that experimental pigs could learn the task reliably, their general level of activity increased, undesirable behavior such as belly-nosing decreased, and they were less fearful in open field and novel object tests. Physiological effects included better immunity and wound healing. Puppe *et al.* (2007) conclude their article by recommending sustained CE as a suitable method for improving the behaviour and welfare of captive animals, and suggesting that designs must be rigorously tested before using on a large scale, with many different species.

Kalb and Puppe (2010) reported their findings regarding long-term cognitive enrichment and its effect on opioid receptor expression in the amygdala of pigs. They kept sixteen male pigs in either an experimental pen or a control pen (N=8 in each group) from seven weeks old, for fourteen weeks until slaughter. The experimental pen contained a 'call-feeding-station' (CFS) that had previously been found by the research group to improve the welfare of pigs, measured using several physiological and behavioural methods (see Manteuffel *et al.* 2009). The pigs were fed via the CFS and had to learn to come

to the CFS upon hearing an individual tone, and to press a button an increasing number of times to receive their ration. The control group was fed the same amount of food from a normal food dispenser twice daily with no cognitive challenge involved. When the pigs were slaughtered at twenty-one weeks the experimental and control groups showed marked differences in brain chemistry in the opioid receptors in the amygdala and hypothalamus. However, there are at least two possible reasons for the differences in brain chemistry other than the CE provided to the experimental group. Firstly, group dynamics were not assessed in any way. Social behaviour is also known to affect the opioid system, as the authors acknowledge, and there may have been some unquantified social variation between the experimental and control group. Cause and effect would be difficult to untangle even with group observation, as any social differences could be caused by either inherent differences between the groups or by the enrichment itself. Secondly while the experimental group received their ration in up to thirty-one small portions over each day, the control group received theirs in just two meals. As having numerous small meals is a more natural feeding rhythm for pigs, and indeed many omnivorous and herbivorous animals, this may have induced the differences rather than the CE. This possibility could be examined by temporally yoking the feeding so that when an experimental animal is fed, a small dose of food is also fed to the control animals. Using organic brain chemistry as a welfare measure is usually prohibitive in terms of cost and animal mortality, at least in the case of zoo and companion animals, so this study provides useful correlates.

Computer-based tasks as CE

Using computer assisted CE is an important step forward as animals have been found to respond well to computer-generated stimuli presented on screens in studies of motivation and cognitive ability (Rumbaugh *et al.*, 1989; Washburn *et al.*, 1989; 1990, 1991; Wilkie *et al.*, 1994; Leighty and Fragaszy, 2003; Wills *et al.*, 2009).

Fagot and Paleressompouille (2009) found that a computerized test system worked very well when training a group of free-living baboons, who presented voluntarily for training. They found the monkeys engaged very often and demonstrated a high learning performance in abstract reasoning tasks.

They suggested that the provision of a computerised learning device could be enriching for captive animals.

'Computer games for animals' are indeed gaining in popularity as enrichment – one of the first studies to investigate the effect of computerised tasks on the welfare of captive animals carried out by Platt and Novak (1997) found that rhesus macaques were able to control a cursor via a joystick to obtain food in their home environment, and showed higher activity levels indicative of improved welfare. They did not seem to habituate to the device over time, possibly because the task varied in difficulty suited to the abilities of the user, helping them to achieve 'flow' and reducing the risk of apathy or anxiety. They also found no evidence of increased aggression. However, Tarou *et al.* (2004) found conflicting results in their study of orang-utans, where subjects were more aggressive and anxious when a computer game device was available due to the increase in social proximity of group members in a predominantly solitary species, and the experience of frustration by displaced individuals, as they found that the joystick was used significantly more by dominant individuals.

Mandell and Sackett (2008) reported on the use of computer touchscreen procedures with infant pigtail monkeys and found that even very young monkeys could use a touchscreen meaningfully and learn operant tasks successfully, suggesting that while there can be problems getting subjects to use a joystick or keys to interact within a computer environment, a touchscreen removes physical barriers from the procedure and enables subjects to engage in cognitive tasks more easily.

Range *et al.* (2008) and Bayer *et al.* (2009) reported the successful use of a touchscreen apparatus with dogs at the University of Vienna, where subjects learned to touch positive stimuli for food rewards in a programme of cognitive experiments. Wilkie *et al.* (1994) and Wills *et al.* (2009) also reported the successful use of touchscreens by pigeons in similar discrimination experiments, while Basile and Hampton (2011) found that rhesus monkeys could complete complex memory tests using a touchscreen. Touchscreen apparatus, with their lack of physical barriers (Mandell and Sackett, 2008) and meaningful use by a variety of species, could therefore provide a means by which flexible, adaptable CE could be presented to captive animals.

Conclusions Regarding CE as a Research Area

Looking at the research examining the effects of CE so far, it appears that CE has the potential to provide challenge, control and even 'flow' to captive animals, it may reduce stereotypic and abnormal behaviour, improve physical condition and cognitive performance and increase positive affect. These possibilities point towards making CE a key research area, and suggest that it is likely that most captive animals would benefit in some way from CE.

2.3.1 Measuring the Welfare Benefits of Environmental Enrichment: Why Measure Welfare?

In many zoo facilities particularly, enrichment programmes are sometimes implemented without any formal recording of (a) the animal's use of the EE, or (b) the effect EE has on the animal in terms of measurable, quantifiable welfare indicators (Fox *et al.*, 2006; Hill and Broom, 2009). It is difficult to extract the cause of welfare improvement after implementing EE if use of the enrichment is not measured. Young (2003) suggests that welfare benefits are not necessarily linked with how long an animal physically interacts with a device. It is also important to note the difference between time available and time used for enrichment items, as this distinction can sometimes be muddled. Studying the ways in which animals interact with EE is also useful in identifying flaws in the design of the EE, for example animals may quickly habituate, the EE may not be fit for purpose, or it may become damaged and dangerous.

2.3.2 How Should Welfare be Measured?

Fraser (2009) argues that historically, Positivism has led to ignoring things that science cannot measure empirically such as animal emotion. This has led in turn to the widespread belief that animals lack the cognitive capacity for emotion, and therefore welfare is of little importance. Behaviour that suggests an animal is experiencing pain has historically been interpreted as little more than an autonomic response to stimuli, and dismissed as unimportant. In recent times, this has been countered by developing empirical methods to access the emotions and experiences of animals, and it is these empirical methods which are gaining respect in the scientific community. Fraser (2009) suggests that careful, methodical qualitative research focusing on

individual differences rather than central tendencies can add an enriching dimension to the more traditional quantitative research techniques.

There are many ways to measure animal welfare, but little consensus as to which methods are the most reliable, or whether methods can be generalized to different species or environments. Azevedo *et al.* (2007) reports that there is wide variation in terms of experimental control and sample size, and a general skew towards positive results. Watanabe (2007) argues that animal welfare does not have a fixed meaning; rather it is a social construct that is flexible and can evolve, which is problematic in terms of operational definition. Techniques differ between laboratory, farm, companion and zoo animals. Young (2003) identified this lack of standardized measurement of welfare as a major flaw in the study of welfare, as it is difficult to compare findings of different studies in turn making it difficult to move forward in making improvements to captive animal welfare. It is proposed here that in order to gain a good insight into animal welfare, and accurately assess benefits caused by interventions, measures looking at several different facets of welfare should be evaluated.

2.3.3 Useful Welfare Measures

Behaviour

The activity budget of captive animals is considered to be one of the most important welfare measures (Young, 2003). If it is evident through behavioural observation that all of a captive animal's natural motivations, or ethological needs, are being fulfilled, then welfare can be said to be optimal. Often, captive animals forage and consume food for only a few minutes per day, whereas in the wild this would take up much of their time – increasing foraging time can be a useful starting point in improving welfare as health problems in captive animals are often caused by an imbalance in nutritional intake versus exercise. Behaviour in the wild is a good base from which to design EE that leads to a more natural activity budget. However, wild animals spending large amounts of time escaping from predators or dominant conspecifics, or surviving through occurrences such as drought or forest fire, would likely consider their welfare to be sub-optimal even though it is natural. Dawkins (1988) also points out that captive animals may manage to adapt well to their environment; and that genetic differences between wild and domesticated animals (e.g. wolves

and dogs, wild cats and domestic cats) mean that differences in behaviour from wild counterparts are not necessarily indicators of lower welfare.

Paul *et al.* (2005) argue that the meaning of behaviours can be misinterpreted, for example an animal may approach a novel object in a positive, exploratory way or in a negative, threatened way. Anticipation is often used as a welfare assessment tool: animals have been found to behave in an increasingly excited way at the signal of a reward if their basic needs have been met (Boissy *et al.* 2007b). However, Van der Harst *et al.* (2003) experimented with two groups of rats, one in enriched housing, the other unenriched, which were trained to expect a reward. The unenriched rats showed more anticipatory behaviour before receiving the reward than the enriched rats which is contrary to the expectation that animals experiencing enriched conditions would have better welfare. In this case the finding may have been due to the enriched rats being more engaged in their home environments than the unenriched rats, or that the experience of receiving a reward was more enriching for the unenriched rats. This demonstrates the possibility of misinterpretation. Behavioural tests such as the elevated plus maze, the open-field test and the hole board (Paul *et al.*, 2005) are often used to measure negative states such as anxiety, for example when testing the effects of drugs, and as they do not require pre-training they are a quick and simple option, but does travel around a maze always indicate exploration or could it be due to an escape attempt due to fear? (Paul *et al.*, 2005).

Held and Spinka (2011) reviewed the usefulness of play (both object and social) as an indication of good animal welfare and found it to be reliable as a measure of emotional valence as animals (both captive and wild) tend to play more when all of their immediate physiological needs have been met, while incidences of play reduce when animals are under stress. They also proposed that play can be 'infectious' and improve the welfare of a group through social transmission.

As Dallaire *et al.* (2012) pointed out as a criticism of EE studies, in order to avoid 'time-filling' or observing the behaviour of subjects during the time they are interacting with EE, behavioural observations should be conducted during times when animals do not have direct access to EE. Recording behaviour while they are interacting with EE means that there is less time for less desirable behaviours such as stereotypic behaviour to be performed, and

incidence of (for example) foraging and locomotion will be artificially raised. The aim is for behaviour to be changed for the better by the presence of EE *at all times*, not just at the time of engagement.

Physical Condition including Glucocorticoids

According to Dawkins (1988), physical health is one of the least controversial and easily measured welfare assays. Newberry (1995) suggests that environmental enrichment should be assessed by the effects it has on an animal's physical health as this can be directly measured. This can be measured in a variety of ways, including visual assessment, heart rate (e.g. Langbein *et al.*, 2004), body weight and condition (e.g. Baumans, 2005), and stress hormone analysis (e.g. Haverbeke *et al.*, 2008).

Visual assessment is easier to carry out on large numbers of animals, such as those in farms or laboratories, or animals that are difficult to handle, but can miss the small changes that bioassays can reveal. However, Fox *et al.* (2006) note that measuring glucocorticoids as indicators of negative states is somewhat problematic as corticoids can indicate excitement as well as stress, therefore missing the valence of the animal's experience. Broom and Johnson (1993) suggest that more than one measure of welfare including physical and behavioural indicators should be taken to get a clearer picture of an animal's well-being, as each technique has its own potential confounds.

Stereotypic Behaviour

Stereotypic behaviour is defined by Mason (1991) as unvarying, repetitive behaviour patterns with no obvious goal or function. Mason and Latham (2004) estimated that over 85 million captive animals perform stereotypic behaviour worldwide. Captive animals are made vulnerable to the development of stereotypic behaviour by their intrinsic motivation to perform behaviours necessary to their survival – that is they are driven to forage, explore, engage in social behaviour, and other species specific drives, and if thwarted by some restriction of captivity, abnormal behaviours such as stereotypic behaviour can develop (Mason, 1991; Boissy *et al.* 2007a). Broom (1983) suggests that as a criterion, more than ten percent of an animal's waking life spent in stereotypic behaviour is an indicator of poor welfare. In operational terms, this seems very high, as an animal that is awake for ten hours per day

would be performing a behaviour that has no functional benefit at all for one hour each day.

There is a possibility that stereotypic behaviour may serve as an adaptation to captivity that improves the welfare of the animal. Mason and Latham (2004) wrote that some animals that engage in stereotypic behaviour show reduced corticosteroid levels compared to animals housed in the same way that do not show stereotypic behaviour, thus raising the possibility that stereotypic behaviour is not as functionless as previously thought, and that actually it provides relief from the stress of captivity. However, the underlying forebrain pathology of stereotypic behaviour in captive animals is similar to humans with schizophrenia and autism (Latham and Mason, 2010), and it follows that, while it may indeed be a coping mechanism for motivations frustrated by captivity, it should certainly not be seen as a positive coping mechanism. Mason *et al.* (2007) call for 'zero tolerance' - stereotypic behaviour in captive animals should simply not occur at all.

There are many reasons why an animal may begin performing stereotypic behaviour, from species differences, to individual differences in genetics, disposition and previous experience, through to the size and quality of housing and contact with conspecifics. Not all animals experiencing the same environment will display stereotypic behaviour to the same extent suggesting a link between individual differences and environmental experience. Latham and Mason (2010) found in an experiment with mice that there was an individual effect on the risk of developing stereotypic behaviour that was linked with a high corticosteroid reaction to environmental change. They found that even sixteen months after an environmental change occurred, the high reaction mice were more likely to be engaging in high levels of stereotypic behaviour. They referred to this as an 'endophenotype'. This suggests that there is likely to be an individual susceptibility to the development of stereotypic behaviour. Ijichi *et al.* (2013) argued that animals prone to develop stereotypic behaviours may have a 'proactive' personality type, which could explain why some captive animals develop stereotypic behaviour whilst others in the same environment do not. The reason for this is that proactivity is also linked with a low frustration threshold and high motivation for reward, and proactive individuals are thought to react to stress by producing more dopamine than less proactive individuals, which is involved in the development of stereotypic behaviour patterns.

Mason *et al.* (2007) linked stereotypic behaviour with increased perseveration, i.e. when an animal is trained in an operant task using reinforcement, then rewards are withheld, an animal that is prone to stereotypic behaviour will show increased perseveration compared with animals that do not show evidence of stereotypic behaviour. They argue that stereotypic behaviour is not linked with reduced learning ability – that animals prone to stereotypic behaviour can learn tasks as well as animals not prone to stereotypic behaviour. However, this is contrary to Hausberger *et al.* (2007) who found that horses that performed stereotypic behaviour were less able to learn an operant task and if they did, took longer to learn it, than horses with no stereotypic behaviour tendencies. They also found that stereotypic behaviour was negatively correlated with resting and proposed that tiredness may account for the reduced learning ability. They also suggested the possibility of neurological deficits.

For primates, particularly those housed in conditions of social isolation, as many are in laboratories, stereotypic behaviour, self-injurious behaviours and other abnormal behaviours are rife, and are correlated with early social deprivation, space restriction, lack of EE and some husbandry procedures (Bloomsith *et al.*, 2007).

Causal factors for stereotypic behaviour are therefore difficult to untangle. Some individuals do indeed seem more prone to developing stereotypic behaviours where other animals experiencing the same environment do not. It is likely that both impoverished raising conditions and long-term housing environment affect proactive individuals with lower learning ability and behavioural inflexibility, pushing them towards developing stereotypic behaviour patterns as a coping mechanism for frustration.

Bloomsith *et al.* (2007) suggest that the simple goal of reducing stereotypic behaviour should be supplemented by assisting the animals to perform in more species-appropriate ways, for example through operant training – to add to their behavioural repertoire rather than just remove a possible coping mechanism. Bloomsith *et al.* (2007) additionally note that moving affected animals into social housing may not be enough to stop negative behaviours once they are established, suggesting that they may be ‘welfare scars’ of past poor welfare. They use a human model of stereotypic behaviour intervention which they argue has been much better-researched than the

analogous animal model. Suggested avenues for research suggested by Bloomsmith *et al.* (2007) include non-contingent reinforcement (reinforcement based only on time rather than behaviour); differential reinforcement of other behaviour (reinforcement based on the subject performing selected other behaviours); functional communication (training an alternative signal to gain attention if that had previously been a reinforcer for negative behaviour); preferred reinforcement (identifying a higher-value reinforcer than the stereotypic behaviour and withholding access if the undesired behaviour is performed); husbandry changes aimed at reducing schedule induced stereotypic behaviour such as more flexible feeding times; and finally increasing the complexity of the environment in which the subject is housed.

Mason *et al.* (2007) and Dallaire *et al.* (2012) agree that EE only slightly reduces stereotypic behaviour in individuals where it is established. Dallaire *et al.* (2012) conducted an experiment with American mink looking at the effect of physical enrichment such as tunnels, objects to manipulate and running water on the severity of their stereotypic behaviour. They hypothesised that stereotypic behaviour reduction should vary according to subjects' level of enrichment use. Their findings were surprising in that while providing EE led to an overall reduction in stereotypic behaviour, the amount of enrichment engagement did not predict the size of stereotypic behaviour reduction. On the contrary, high users of enrichment showed the smallest reduction in stereotypic behaviour. However, high users also had the lowest starting levels of stereotypic behaviour, suggesting there is an interaction between the likelihood to begin performing stereotypic behaviour patterns and the likelihood to engage with EE. One criticism of this study is that it is reported that mink that performed higher levels of stereotypic behaviour did rest in the new EE tunnels more often than in their usual preferred spots, which could actually be considered a use of the enrichment if the time spent in physical contact with the EE had been measured instead. This suggests that 'enrichment use' needs to be carefully quantified.

Researchers providing animals with EE should also consider the effect that removing EE after experiments will have on the animals – Latham and Mason (2010) found that mice that were moved from enriched to unenriched conditions showed a higher level of stereotypic behaviour than mice raised in unenriched conditions, while starlings (Bateson and Matheson, 2007) and pigs

(Douglas *et al.*, 2012) experiencing barren conditions after being housed in an enriched environment made more negative judgements about ambiguous stimuli than starlings and pigs experiencing the opposite environmental manipulation. This suggests that captive animals suffer more when their environment becomes less enriched, and this is therefore an important consideration for all captive animals, but perhaps particularly for animals captured from the wild and housed in laboratories and zoos.

Stereotypic behaviour seems therefore to be a useful indicator of poor welfare in captive animals, and a primary aim of any intervention should be to prevent, eliminate or at least reduce the performance of stereotypic behaviour by captive animals.

2.3.4 The Link Between Operant Response and Emotion

The possibility that animals can experience emotions, whether they are at all analogous to human emotions and whether animals are consciously aware of their emotional experience is subject to much controversy. Panksepp (1998) stated that all mammals have the necessary brain structure for feeling emotions, although his argument has been criticized for having a lack of evidence (e.g. Barrett, 2006a; Barrett *et al.*, 2007). Panksepp (2011) followed this up by acknowledging resistance within the scientific community to the concept of animal emotions, but argues that recent advances in neuroscience have allowed objective investigation into cross-species emotion (e.g. Panksepp and Lahvis, 2011) and evidence is mounting for a 'Law of Affect' where emotions may shape brain structure in order to learn from experience and improve fitness. Very recently, Leliveld *et al.* (2013) have reviewed the evidence for a universal lateralisation of affect – this means that of the many animal brain studies they reviewed, the structures that process positive and negative emotions seem to be separated in opposite brain hemispheres, with positive emotions being processed in the left hemisphere while negative emotions are processed in the right hemisphere. This further supports the notion that animals do experience emotions of different valence similarly to us and each other. Also see Paul *et al.* (2005) for an in-depth discussion of animal emotion and affect.

In 1995, Carey and Fry published a research paper looking at the possibility of 'asking' pigs how anxious they felt. This was achieved by training pigs on an operant task then conditioning them to respond in a certain way

when they were administered with a drug with known anxiogenic effects, which the pigs then generalized to other anxiety-provoking situations such as changes to their environment and exposure to unknown conspecifics. This study provides evidence that animals can actually tell us if they are experiencing anxiety, and perhaps other emotional states, through their operant responses. This achievement is conceptually hugely important for animal welfare, although it would generally be difficult and unethical to implement as it involves the use of drugs known to reliably induce a state of anxiety in order to train the animals that what they are feeling is in fact anxiety. Alternatively, the use of (for example) environmental changes correlated with other measures of reduced welfare for training may not be equally anxiety-provoking in all individuals, and therefore when they 'tell' us they are feeling anxious, they may be feeling excited, bored, interested or other emotions instead.

In another psychopharmacological welfare study, Sherwin and Olsson (2004) found that mice would self-administer less of an anxiolytic drug when housed in enriched cages compared with mice in standard laboratory cages, suggesting that the standard-housed mice experience more anxiety which is relieved by the drug.

This kind of research is useful for providing quantifiable, reliable and insightful findings into animal emotion. However, in zoos particularly, conducting psychopharmacological research may be difficult as the visiting public may consider it unethical.

Preference and motivation tests (see Kirkden and Pajor, 2006 for a review) seek to quantify the value of various resources presented to animals such as enrichment items, substrates, foods and conspecifics. This can be very useful in discovering what animals actually find important. In preference tests, animals are given several choices and their preferred item/location or conspecific is noted. When deciding which rewards to give in operant studies, a favourite reward can help as an incentive. In motivation tests, which use economic theory as a basis (Lea, 1978), animals are given restricted access to something, and the effort they are willing to put in, i.e. effort in door pushing, number of lever presses, is recorded and compared with the effort made for other resources. Some things, like food or water, are essential resources for survival, for which animals show 'inelastic' demand (they will pay more as the cost goes up, to a point), whereas some resources elicit more 'elastic' demand

meaning that a rising cost can quickly become unacceptable, and this varies by species as well as between and within individuals. This variable measure of the elasticity of demand can serve as an indication of the need for certain enrichments, giving us an insight into what is really important for animals (Dawkins, 1983; Manser *et al.* 1996; Sherwin and Nicol, 1997; Olsson and Keeling, 2002). For example, farmed mink will pay large operant costs to access a water bath for even a short time, and they have higher levels of cortisol if then prevented from doing so, suggesting that for mink, access to a water bath is particularly important for their welfare (Mason *et al.*, 2001).

Cognitive Bias

A new development in the study of animal welfare has come from research on depressed and anxious humans. Humans experiencing a positive affective state tend to interpret ambiguous stimuli in a more positive way, i.e. they are more optimistic than people experiencing a negative affective state such as depression or anxiety (Eysenck *et al.*, 1991). This can be used to assess mental state in a non-verbal way. Boissy *et al.* (2007a) identify cognitive processes as having a key role in the perception of events and situations by animals, which in turn affects and is affected by their emotional state. They suggest that both cognition and emotion must be considered when assessing animal welfare. Mendl *et al.* (2009) reason if you can accurately assess this two-way flow between cognitive processes and emotional state in humans, it is possible that the same phenomenon exists in other species and can be assessed in a similar way. Researchers are now attempting to investigate whether animals have a similar cognitive bias to humans in their judgement of ambiguous stimuli, which may in turn shed light on their affective state (e.g. Harding *et al.* 2004; Mendl and Paul, 2004; Paul *et al.*, 2005; Bateson and Matheson, 2007; Matheson *et al.*, 2008; Mendl *et al.*, 2009; Brilot *et al.*, 2010; Bateson *et al.*, 2011; Bethell *et al.*, 2012; Douglas *et al.*, 2012, Starling, 2012; Briefer and McElligott, 2013). The study of cognitive bias in animals has gained popularity as an animal welfare assessment tool in the field of EE as it seems to be quite a sensitive technique that varies reliably according to experimental manipulations (Harding *et al.*, 2004; Mendl and Paul, 2004; Paul *et al.*, 2005; Burman *et al.*, 2009), and it has been supported by other established welfare measures (*behaviour*: Mendl *et al.*, 2010; *biochemical analysis*: Bateson *et al.*,

2011; Sanger *et al.*, 2011; *stereotypic behaviour*: Asher *et al.*, 2009; Brilot *et al.*, 2010; Pomerantz *et al.*, 2012). A grant has been recently approved by The National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) for Professor Mendl and colleagues at The University of Bristol running from 2013-2015 for a project entitled 'Development and validation of an automated test of animal affect and welfare for laboratory rodents'. Mendl and colleagues propose to thoroughly test the validity of the cognitive bias paradigm as a welfare measure for mice, and develop an automated device to make such testing easier, quicker and more applicable to large numbers of animals.

An important part of this development of an objective way to collect evidence of animal emotion and affect is that it allows us to rectify the skew towards collecting evidence only of the reduction of negative welfare markers (Boissy *et al.* 2007b), such as reduced stereotypic behaviour, in EE applications. We can see instead if animals are more optimistic, rather than just less pessimistic.

The basic paradigm for cognitive bias study involves the measurement of positive and negative value judgements of ambiguity by subjects, usually by training subjects on an operant task where S+ and S- are on a continuum, for example a colour spectrum (dark grey versus light grey stimuli), space (left versus right) or tone (low versus high). Subjects are then offered several ambiguous, intermediate 'probe' choices, such as shades of grey, different spatial positions or medium tones. If a subject tends to interpret ambiguous stimuli in the same way as S+, it suggests they are more optimistic, while more pessimistic subjects tend to interpret ambiguous stimuli as S-.

Bateson and Matheson (2007) investigated the effects of environmental enrichment on the optimism of captive starlings using a cognitive bias paradigm. They trained the birds using a white/dark grey, go/no-go task with stimuli as lids covering food rewards (S+ palatable, S- unpalatable), then presented stimuli of ambiguous grey shading, to see if the starlings would interpret them as positive or negative stimuli, enabling subject's classification as 'optimistic' or 'pessimistic'. The results of the study were that starlings exposed to enriched conditions, and then moved to unenriched conditions were significantly more pessimistic in their interpretation of ambiguous stimuli than starlings that experienced the conditions in the opposite order, suggesting that moving from

an enriched environment to a barren one results in poorer welfare for starlings. The starlings were trained and tested in their home cages, which removes any confounds associated with moving such as increased stress from handling and as such is better for welfare. Lid-flipping as an operant task is a naturalistic foraging behaviour corresponding to flipping dead leaves for a starling and is therefore quicker and easier to train than other paradigms such as the one used by Harding *et al.* (2004) involving tone discrimination coupled with lever pressing. However, training a go/no-go task is arguably more difficult than tasks where there is always a 'go' option such as a simultaneous discrimination as there is a reliance upon the subject choosing to do nothing, which must be differentiated from not being motivated to do anything, as this may be caused by something other than interpreting the stimulus as S-, for example lack of feeding motivation due to satiation or nausea. There was no counterbalancing between shades in the study as all subjects had white as positive and dark grey as negative. To remove any confounds related to visual acuity stimuli should be counterbalanced between subjects – if for example the darker stimulus was easier to discern from the background than the lighter stimulus for subjects, therefore drawing attention quicker or more reliably to that stimulus, the whole experiment would be skewed towards selecting that stimulus more often. Using no reward and/or a time-out for S- may be preferable to using an aversive nausea-inducing reward for S- as this may put subjects off from taking any risk at all with ambiguous stimuli. Finally, the task had non-reinforced probes so there was a bias towards negative responses from animals learning that probes are as bad or worse than negative stimuli, especially when there are repeated measures. In designs where there is counterbalancing between experimental manipulations (as in the study by Bateson and Matheson, 2007), this effect is at least balanced, but the whole experiment is skewed towards a more negative interpretation, so perhaps all subjects seemed less optimistic than they really were. This is further supported by fact that only the ambiguous stimuli 20% grey was significant in the cage order interaction (cage type x order), and with only a trend towards significance in cage type.

Briefer and McElligott (2013) found the cognitive bias paradigm effective in assessing the optimism of rescued goats that had previously suffered neglect. At a rescue centre where goats received good care, goats who had previously suffered neglect were either more optimistic, or at least as optimistic,

as control goats who had never suffered neglect, suggesting that goats can recover from bad treatment. This opposes the theory of 'welfare scars' left over from previous poor welfare experiences, as indicated by stereotypic behaviour and other welfare measures (Bloomsmithe *et al.*, 2007). However, each group (neglected, non-neglected) was further divided for analysis into males and females, and it is only the females in the neglected group (N=4) that showed a more optimistic bias than females (N=5) from the non-neglected group. Although statistically the finding was significant, there are two issues with accepting this finding without further investigation. Firstly, the overall sample size is very small. This can lead to confounds within the sample, which my own experiment two demonstrates: one group of subjects were trending towards higher optimism before any experimental intervention, despite being randomly selected for that group. This leads to the second criticism, that the division of goats into poor and control welfare groups was based on some questionable decisions – a goat that is obese (poor welfare group) may not subjectively experience lower welfare than a goat kept on an optimum diet, while a control group male was described as 'unwanted, kept escaping', suggesting that its home environment may not have been desirable. Another control group male was included although its companion had died, which could be considered a reason for poor welfare. A possible remedy for this would have been to assess welfare objectively when the goats first arrived at the shelter, including gathering cognitive bias data. This would have provided a baseline measure and given some weight to the classification of groups as having previously experienced good and poor welfare. While the authors reached a valid conclusion based on the data, it may be improved by having a clearer definition of 'poor' and 'control' welfare and including more subjects in order to validate the findings.

Cognitive bias as a measure of animal welfare seems to be reaching interesting conclusions and providing a reliable insight into animal emotion that was previously based upon contested subjective opinion. However, as it is a new technique, there are many variations in methodology, each with their own strengths and weaknesses, which make the results difficult to compare. Strict experimental design and methodology including repeated baseline measures, large enough sample sizes, counterbalancing, experimental group homogeneity

or heterogeneity and other techniques should be used to reduce confounds. In addition, three fundamental issues need to be resolved, as discussed below.

The 'Advance Key' procedure

Matheson *et al.* (2008) criticize the go/no-go procedure used in some studies (e.g. Bateson and Matheson, 2007) for being insensitive to the possibility that other factors may lead to a general reduction in movement and drive to seek reward, rather than pessimistic judgement *per se*, therefore skewing the results in favour of a pessimistic judgement bias. If the S- is not too aversive, subjects may not be sufficiently discouraged to sit out a trial (no-go), leading to a slow learning rate and an overinflated optimism score when evaluating cognitive bias, as all of the probes may be selected regardless of valence. A possible solution to this is the active choice method, used by Matheson *et al.* (2008), Enkel *et al.* (2010), Brydges *et al.* (2010), Anderson *et al.* (2012) and Anderson *et al.* (2013) among others, whereby a training paradigm involving an active response for a positive stimulus (rewarded), and a different active response for the negative one (to avoid punishment, to gain a smaller reward, or gain a reward associated with a delay). But what if the subject interprets the ambiguous stimulus as positive, but is unwilling to run the risk of possible punishment? This may lead to an increase in non-response or an increase in interpretation of probes as negative, in order to minimise risk, as rats and humans have been found to prefer avoiding punishment to gaining a reward if the stimuli are ambiguous (Harding *et al.*, 2004; Enkel *et al.*, 2010; Anderson *et al.*, 2012). What is needed is a neutral option for subjects to choose in each trial, which may accelerate learning and provide an outlet for the motivation to participate. Leyland and Honig (1975) outlined a procedure which enabled pigeons being trained on a discrimination task to skip to the next trial with no other positive or negative consequence, which they described as an 'Advance Key Procedure' (AK). Leyland and Honig (1975) suggested that pigeons can quickly learn to make the AK choice in the presence of S- rather than S+, indicating that it is a discriminated stimulus. So, if the training stimuli are shown along with an AK (e.g. a different symbol in a separate location for clarity) as well as with the probe stimuli in the testing phase, the advance key provides an opportunity for subjects to always respond (i.e. go/go instead of

go/no-go), thus removing the chance that some no-go's are due to lack of attention/apathy rather than an active choice, and subjects learn that the AK holds no negative associations but also no reward, that it is neutral except for the property of moving the trials along so they may get a positive option next time, then this clarifies the intention behind their response.

The advance key procedure is therefore advantageous compared to the positive/negative active response procedure because it allows for a third category when examining ambiguous probes, i.e. they can be interpreted as not positive or negative, but ambiguous, as the advance key is not associated with a reward or a punishment, and the subject can then choose that they would rather not get a reward than risk a punishment. Therefore, if they are interpreting a probe as a positive stimulus, they would choose it over the advance key. The availability of an advance key is therefore an ethical improvement on just having a positive and negative option, as subjects can always avoid a punishment, however innocuous the punishment may be.

In summary, in a go/no-go operant task design, there is reliance upon subjects to be able to do nothing. In essence you cannot really tell the difference between a subject responding 'no' or 'not sure' or 'not interested'. In choice tasks there is a chance that the subject would prefer to not receive a reward than risk a punishment. The AK option provides a true 'yes/no' alternative as when S+ is paired with AK the correct answer is always S+, and when S- is paired with AK the correct answer is always AK. When testing responses to ambiguous stimuli the addition of AK tells us if subjects are interpreting the ambiguous probe stimulus as S+ or S-.

Reinforcement of Ambiguous Probe Trials

Most cognitive bias studies using animals have been carried out to examine the effect that environmental improvements have on welfare, requiring the use of repeated measures. When probe stimuli are all given the same 100% reinforcement, (i.e. in most current paradigms all of the probe stimuli are unrewarded), the probability of probe stimuli eliciting a reward drops to zero. The fundamental issue here is that all generalisation studies, where stimuli A and Z are trained then the responses to B,C,D etc. are tested, encounter the problem of probe reinforcement. Rewarding all probes as if they were S+ is

problematic as subjects will learn to interpret them all as S+, while never rewarding the probes leads to subjects learning to interpret them all as S-.

In most cognitive bias experiments reported thus far, subjects are presented repeatedly with unreinforced ambiguous probe trials (e.g. Bateson and Matheson, 2007; Burman *et al.*, 2008; Starling, 2012), which does skew the responses in favour of pessimism over trials as the subjects begin to learn that probe trials are unrewarded and therefore probe stimuli are effectively S-. This suggests that the experimental cognitive bias paradigm as it stands should not be used to measure the same subjects over time, although it is becoming a common technique used to assess welfare improvements due to EE.

An example of this problem can be seen in the study by Burman *et al.* (2008). Ambiguous probes were baited with inaccessible rewards (to control for olfactory cues) in their study of rat cognitive bias, stating that this 'minimised any undesirable associations between the ambiguous locations and the reward outcomes that may have been learned rapidly if the ambiguous probes had been rewarded' (Burman *et al.* 2008 p. 805). If rewarded (positive) and unrewarded (negative) stimuli are on opposite ends of a spectrum, and it is expected that rewards enhance the speed of learning the positive association, it follows that never rewarding would also increase the speed of learning the negative association. The probes are supposed to be ambiguous; therefore the possibility that they *maybe* rewarded should enhance their ambiguity.

This key issue was raised by Doyle *et al.* (2010) who found with sheep that repeated testing of ambiguous unreinforced probes over several weeks led to a significant drop in response to the probes, despite introducing no experimental manipulations consistent with lower welfare. Unless a study is properly counterbalanced this could lead to the erroneous interpretation of lower welfare over time. Even in properly counterbalanced experiments, this downturn over time could skew the interpretation of all subjects' responses erroneously towards pessimism.

Brilot *et al.* (2010) also acknowledged this problem with using cognitive bias tasks for assessing optimism in repeated measures experiments, and through reanalysis of the data in Bateson and Matheson (2007) found that subjects learning that the ambiguous probes were unrewarded had in fact affected the proper interpretation of subjects' optimism scores adversely. Brilot *et al.* (2010) suggest that the usual cognitive bias paradigm of training S+/S-

then presenting unreinforced probes should not be used in repeated measures design experiments, and should instead only be used for experiments using single-exposure probe trials to prevent subjects learning that probes are never rewarded.

Conversely, Franks (2012) used an ambiguous maze to look for welfare differences in rats and positively reinforced all of the ambiguous maze arms which were then used in a repeated measures study of enrichment, which may have resulted in a possible type II error as subjects became indistinguishable from controls over time. Franks (2012) draws attention to this when discussing her experiment, and states that further research is needed regarding repeated presentation of ambiguous options.

Another related problem, seen in Matheson *et al.* (2008), was that the possibility of reward in some designs can drop from 1.0 (always one rewarded option) during training to a lower proportion during the probe sessions caused by the lower total number of S+ available (e.g. in Matheson *et al.*, 2008, the possibility of reward dropped to 0.33) which may reduce the welfare of subjects, creating a possible confound in experimental interpretation.

One solution to this is to reward the S+ on an intermittent schedule so that occasional unrewarded probes do not lead to extinction, however if S+ is trained on an intermittent schedule it can lead to slow acquisition. Brilot *et al.* (2010) used high and low value rewards as S++ and S+ rather than the more usual S+ and S- and attempted to slow extinction due to ambiguous probes being unrewarded whilst speeding up initial training by training the stimuli on a 100% reward schedule then gradually reducing the rate of reinforcement of the training stimuli to the level experienced in the probe trials, as probe trials were never rewarded. However, they found that latency to choose stimuli increased steadily, indicating that the starlings learnt that the ambiguous probes were never rewarded. Meanwhile, Brydges *et al.* (2010) used a similar S++/S+ technique and also attempted to address this problem by not rewarding twenty-five percent of S++ and S+ training trials. Again, probe trials were never rewarded, thus making the ambiguous probe a less favourable option than either the S++ or S+ training stimuli. In their experiment, Brydges *et al.* (2010) found that having unreinforced probe trials did not lead to a downturn in responses, suggesting that extinction was not taking place, however subjects would presumably eventually learn that ambiguous probes were effectively

worse than both S++ and S+ training stimuli if they were never rewarded, therefore this technique is probably more suited to experiments with few repeated experimental conditions.

Anderson *et al.* (2012) and Anderson *et al.* (2013) rewarded ambiguous probes according to which training stimulus they were closest to in type. This means that half the probe types were associated with reward and half the probe types were associated with punishment. However, as each probe was then associated with the same level of reward each time, subjects could quickly learn the association in the same way they learnt S+/S-. The experiments were not repeated with the same participants, but ambiguity would presumably diminish in the same way that never rewarding or punishing ambiguous stimuli does if the technique was used in a repeated measures experiment.

An important question to answer when developing cognitive bias tasks is whether we are designing games based on ambiguity or risk. Whilst risk-taking makes use of known probabilities, ambiguity means that probabilities are unknown. Arguably, any cognitive bias task paradigm that involves more than a single presentation of a probe stimulus gives the subject the opportunity to base their future behaviour when presented with the same stimulus on the probability of a certain outcome, meaning that they learn through experience. The aim here is not to completely remove any chance of learning, but to slow it to the point at which responses to probes remain useful over repeated presentations.

Another possibility is to reinforce probes probabilistically, where probes are reinforced according to a predefined schedule, for example reinforcing each probe once every four times it is shown. This brings in the complication that subjects will begin to learn the probability of each probe being rewarded although learning will arguably be slower than on a simpler schedule.

It is therefore proposed that a 0.25 pseudo-randomised reinforcement schedule for the probe stimuli may work to elicit the underlying cognitive bias without leading to extinction. The reason for this proposal is that previous experiments (e.g. Wills *et al.*, 2009) designed to assess the classification of stimuli by pigeons, humans and squirrels, in the same laboratory setup with similar subjects, set-up, stimuli, reinforcement schedule and experimental aims showed behavioural variability rather than extinction, suggesting subjects did not learn to always associate the probe stimuli with either S+ or S-, but interpreted them as ambiguous, thus enabling researchers to establish the

mechanics of visual classification of compound stimuli by pigeons, humans and squirrels.

Cognitive Bias Training as Enrichment

Operant discrimination is one of the proposed cornerstones of CE. As operant training has been found to be beneficial for welfare (e.g. Pomerantz and Terkel, 2009) there is a possibility that the operant training for a cognitive bias task may actually confound the welfare changes it is used to measure by raising welfare. This raises questions about the validity of the technique as an independent measure, and of the EE it is used to test. If conditions are properly balanced, any effect of the cognitive bias task training should be equal throughout experimental groups. However, in order to have confidence in findings, the effect of the cognitive bias task itself on welfare should be measured, perhaps by using naturalistic experiments based on untrained behaviour, using ambiguous stimuli such as natural food items dyed different shades on a continuum, mimicking higher and lower quality food items. More optimistic animals should investigate any potential food source, whereas more pessimistic animals would be expected to only choose the normal looking food items. This naturalistic experiment could be run before, during and after the training for the operant cognitive bias task, as well as during any experimental EE interventions.

3.0 Contribution of this Thesis

The research described in this thesis makes a contribution to knowledge in this field by examining the effect of CE on the welfare of captive laboratory and companion animals in a methodical and standardized way, enabling comparisons between studies, and improvements to captive animal welfare to be made. This is a popular research area and there have been many advances in recent years particularly with zoo and farm mammals. However to date not much is known about the effect of CE upon laboratory-housed birds or companion animals such as dogs; the effect of group versus individual housing on the usefulness of CE; and CE has not previously been measured by using the matrix of measures used here. In addition to this, I have also suggested several improvements to the cognitive bias paradigm and designed an operant device for providing CE and measuring welfare through cognitive bias.

Swaigood and Shepherdson (2005) called for improvements to be made in EE research in order for the information to be useful in improving the welfare of captive animals. In particular, they suggested an adequate sample size, a repeated measures experimental design, clear description of EE and procedures and the provision of full statistical information about the behavioural changes observed during the experiment, including standard error. I followed these guidelines in order for this research to be put to use in improving the welfare of captive animals in all settings. To do this I conducted a series of experiments designed to look at the importance of CE for the welfare of group- and individually-housed social animals: pigeons (*Columba livia*) and domestic dogs (*Canis familiaris*). The following questions will be answered after presenting my findings:

Questions to be Addressed by this Thesis

1. What is CE, and does it improve captive animal welfare?
2. Is typical, feasible CE more beneficial to captive animals than typical, feasible PE?
3. Is increasing environmental cognitive complexity associated with increasing welfare benefits?
4. Is CE better delivered in a group or individually accessed setting?
5. What affects the welfare of individually kennelled dogs and can their welfare be improved by CE?
6. Is it possible to develop an automated CE device that prevents habituation and is easy to implement in many laboratory, zoo and companion animal settings (e.g. kennel, home environment) that improves (and also measures) welfare?
7. Is stereotypic behaviour a good measure of welfare and can it be reduced through CE?
8. Does optimism increase in captive animals when they are provided with CE?
9. Does the cognitive bias measurement paradigm suit this kind of research, what are the current weaknesses and how can they be improved upon?

4.0 Introduction to Experiments One and Two

The next chapter discusses the two experiments carried out with a group of sixteen captive laboratory pigeons. CE has been found to be beneficial to the welfare of farm (Manteuffel *et al.*, 2009) and zoo (Clark, 2011) animals. Major criticisms of most environmental enrichment (EE) implemented in laboratory, farm and zoo animal facilities include the fact that it is usually based upon physical items that are easily habituated to, do not reduce stereotypic behaviour sufficiently and are not measured in a consistent, methodical, objective way. My first experiment compared the effects of a typical physical enrichment (PE) manipulation to a simple cognitive enrichment (CE) manipulation by measuring baseline and experimental phase physical condition, behaviour and optimism via a cognitive bias paradigm similar to that used by Bateson and Matheson (2007). CE was found to have significant positive effects upon the welfare of the pigeons, particularly by reducing stereotypic behaviour and increasing optimism. The finding that CE improved welfare for this group of captive pigeons raised two further questions: firstly was it the cognitive challenge of the CE device, or the increased time it took to interact with physically that made it more enriching? Secondly, what effect did presenting enrichment to the group as a whole have on the enrichment value of it for individuals? For the second experiment I therefore compared two versions of the CE device which differed only in the complexity of the puzzle they presented, and I divided the group of pigeons into two and provided access to one of the two devices individually each day to allow for individual learning and protection from dominant individuals. I found that, again, the provision of CE led to welfare improvements in this group of pigeons, but there seemed to be no difference in effect between the low- and high-complexity CE devices apart from that the pigeons used the high-complexity CE device more over the enrichment phase. Whilst my study did not find any welfare correlates with this, it is possible that over time, engagement with and consequent welfare benefits from the low-complexity device would reduce while engagement with and welfare benefits from the high-complexity device would continue. Group and individual presentation of CE seems equally beneficial in this case but may not be so simple in the case of more aggressive animals. Learning ability and dominance were negatively correlated with each other but not with any of the welfare measures in this

study. Optimism as a welfare measure worked reliably, with one change from the usual cognitive bias paradigm (Bateson and Matheson, 2007), the ambiguous probes were rewarded on a variable interval schedule rather than remaining unrewarded through repeated trials, and this was found to maintain an ambiguous interpretation. Through these two studies I began to develop a useable CE method and welfare measurement system.

4.1 Chapter Two: Experiment One and Two

A complex cognitive enrichment procedure improves the welfare of group-housed pigeons to a greater extent than a physical enrichment procedure, and group presentation is preferred to individual enrichment.

Louise N. Millar*, Stephen E.G. Lea, Lisa A. Leaver

**Corresponding author: Louise N. Millar, University of Exeter, School of Psychology, Washington Singer Laboratories, Streatham Campus, Perry Road, EX4 4QG. Email:*

L.N.Millar@exeter.ac.uk.

Abstract

The provision of suitable, effective environmental enrichment (EE) for captive animals is an important issue in animal welfare, as it has been shown to improve the neurobiology, physiology and behaviour of captive animals. EE is usually limited to inanimate objects which require physical manipulation, known as physical enrichment (PE). However, cognitive enrichment (CE), or enrichment which requires animals to use their cognitive abilities, is thought to have significant effects on captive animal welfare (Meehan and Mench, 2007; Manteuffel *et al.* 2009). This is an important consideration in the use of animals for cognitive research. I designed an experiment to compare the effects of CE and PE on laboratory animal welfare. Using a laboratory-based, group-housed population of pigeons (*Columba livia*), a species extensively used in cognitive experimentation, I measured welfare based on activity budget changes, stereotypic behaviour, physical condition, 'optimism' (Bateson and Matheson, 2007) and use of enrichment during two ten-day experimental phases where (a) a cognitive enrichment device (CED) and (b) a physical enrichment device (PED) were introduced to the group of pigeons for one hour a day, with baseline measures taken before, during and after each experimental phase, to examine how welfare was affected.

Key findings were that during the CE phase, the pigeons displayed significantly less stereotypic behaviour, and were significantly more 'optimistic', than during the baseline phases and the PE phase. Optimism and stereotypic

behaviour were not correlated within subjects suggesting that different subjects expressed their experience of poor welfare differently. Subjects also used the CED significantly more than the PED, with higher enrichment use predicting a larger increase in optimism.

In order to find out why subjects used the CED more, a second experiment was devised to test the enrichment value of high- and low-complexity CE devices using a laboratory-based, group-housed population of pigeons (*Columba livia*), a species used extensively in cognitive experimentation. I measured welfare based on stereotypic behaviour, 'optimism' (Bateson and Matheson, 2007) and engagement with enrichment during a ten-day experimental phase where two groups of pigeons were given individual access to either (a) a low-complexity cognitive enrichment device (LCCED) or (b) a high-complexity cognitive enrichment device (HCCED) for ten minutes per day.

During the enrichment phase, both groups of pigeons displayed significantly less stereotypic behaviour, and were more 'optimistic', than during the baseline phases. Stereotypic behaviour and optimism were not correlated within individuals suggesting that there was variation in how individuals experienced and expressed poor welfare. Higher use of enrichment predicted a larger improvement in optimism. Over time both groups used the enrichment devices more, although the high-complexity device was used more overall, with little evidence of habituation. Group and individual presentation of enrichment both improved welfare, but subjects did not use the individually presented enrichment until day five, and the relatively stable dominance hierarchy was disrupted by the daily upheaval brought about by individual training and enrichment. Dominance was negatively correlated with learning ability, perhaps due to social pressures caused by the disturbance. It appeared therefore that a higher-complexity cognitive enrichment device presented to the group was the most beneficial method of enrichment overall in this case.

Keywords: Environmental enrichment, captivity, animal welfare, cognitive bias, cognitive enrichment, stereotypic behaviour, pigeons (Columba livia).

4.2 Introduction

4.2.1 Environmental Enrichment (EE)

Improvements to the environment of captive animals going beyond the most basic provisions of food, water and space to move around are known as environmental enrichment. The requirement that animals should be able to perform their full range of natural behavior (UK Animal Welfare Act, 2006) is very difficult for animal keepers to comply with in a captive setting. Bloomsmith *et al.* (2007) and Shuangying *et al.* (2009) argue that the captive environment can be considered damaging by its very nature, which is typically unstimulating, unchanging and unchallenging. Fox *et al.* (2006) propose in their review that effective EE for captive animals such as rats influences their neurobiology, physiology (shown by better body condition, more resistant immune systems, and faster recovery from illness) and behaviour (shown by reduced anxiety and improved cognitive function and less fearfulness and hyperreactivity in new situations). In order to survive in the wild, animals must use their cognitive skills, such as navigation, memory, perception, learning, recognition, categorization, problem-solving, decision-making, reasoning, as well as sometimes very complex social cognition such as living in a hierarchy and pack-hunting. It could be argued that an animal that never has to make any efforts to ensure its survival through cognitive effort does not have good welfare.

Meehan and Mench (2007) wrote that the current trend for environmental enrichment (EE) is to satisfy only the physical needs of an animal, for example through scatter-feeding to increase the time spent foraging, providing substrates or structural improvements, known as physical enrichment (PE). They propose that because of the over-reliance upon PE, captive animals are not challenged beyond their basic perceptual and motor skills, and suggest that there is a need for research into the effects of cognitive enrichment (CE) on the welfare of captive animals. Manteuffel *et al.* (2009) have found that most enrichment given to farm animals is based on inanimate appetitive or structural applications which have the disadvantage of being habituated to very quickly.

4.2.2 Cognitive Enrichment (CE)

Milgram (2003) originally coined the term 'cognitive enrichment' for animals in a research article describing a battery of cognitive tasks given to

dogs, found to slow age-dependent cognitive decline. Sale *et al.* (2009) reviewed the literature concerning EE and brain function, and found that engaging with complex cognitive tasks has the capacity to modify behaviour, enhance learning and memory, and reduce ageing-related cognitive decline in several mammal species including humans. Manteuffel *et al.* (2009) define CE as any type of enrichment that requires cognitive effort from the animal, leading to an ability to exert active control of the environment. Appropriate CE, in the form of solvable problems or tasks, has the potential to be enriching, as captive animals can develop the cognitive skills they need to progress and master the tasks, leading to increased welfare as the animal learns to successfully cope with a cognitive challenge (Franks, 2012). Wechsler and Lea (2007) suggested providing captive animals with such cognitive tasks to reduce boredom in captive situations, thus improving welfare.

Puppe *et al.* (2007) examined the effect of a novel cognitive task given to growing pigs, in this case learning acoustic cues to signal access to different feeding-stations. They found that experimental pigs could learn the task reliably, their general level of activity increased, undesirable behaviour decreased, and they were less fearful in open field and novel object tests. Physiological effects included better immunity and wound healing. Puppe *et al.* (2007) conclude their article by recommending sustained CE as a suitable method for improving the behaviour and welfare of captive animals, and suggesting that designs must be rigorously tested to establish their usefulness before using on a large scale, with many different species, rather than assuming that if benefits are found with one group of animals, they will be seen in all animals.

Manteuffel *et al.* (2009) describe in their review of CE techniques for farm animals how successfully coping with a cognitive challenge has been found in pigs to be linked with increased dopamine production, and also suggest that stereotypic behaviour may be reduced by the introduction of CE, due to the animals being less bored and frustrated by their environment. Langbein *et al.* (2009) suggest that giving captive animals the opportunity to use their cognitive abilities by presenting them with CE, and allowing them to control their environment to some extent, should be considered of great importance in increasing their psychological well-being. Hagen and Broom (2004) found that cattle given a cognitive task showed emotional reactions to their own learning, behaving more excitedly when they improved upon a task. Langbein *et al.*

(2009) conducted a study with group-housed dwarf goats, examining the possibility that goats seek out and benefit from cognitive challenge. Their study consisted of training the goats to use an 'automated learning device', where correct responses to stimuli were rewarded by sips of water. They found that the goats continued to use the automated learning device even when water was available freely elsewhere (a phenomenon known as contrafreeloading, first systematically demonstrated by Neuringer, 1969), and suggest that this is evidence that the goats found cognitive challenge intrinsically rewarding.

If CE improves captive animal welfare, special consideration should be given to animals that are used in cognitive experiments: if animals kept in barren conditions that affect their cognitive abilities are then used in cognitive experiments, it is likely that those animals will provide lower-grade results than animals kept in cognitively stimulating environments. The use of laboratory animals in cognitive experimentation is based on the assumptions that (a) the animals are fit to work under the conditions they are kept, and (b) they will provide reliable, valid results. Neither assumption is justified if animals are kept in sub-standard conditions. In addition, if CE is beneficial to laboratory animals used in cognitive research, there is the possibility that far from lowering the welfare of the animals, the experiments themselves are increasing the welfare of the animals used. This is an important consideration in the use of animals for cognitive research.

The reports described above indicate that so far, in the relatively new area of CE and its effects on animal welfare, CE seems to be a promising avenue for development as it has so far been shown to be effective in improving the welfare of farm animals (e.g. Manteuffel *et al.*, 2009). What needed to be examined is the effect of CE compared with other more typical methods of EE such as physical enrichment (PE), as well as measuring the effects of CE upon captive animal welfare in a way that promotes easy comparisons between CE devices, species and settings. Experiment one, described herein aimed to bring together these factors, in a laboratory setting, with pigeons, a species commonly used in cognitive experiments.

4.2.3 Measuring Animal Welfare

There are many ways to measure animal welfare, but little consensus as to which methods are the most reliable, or whether methods can be generalized

to different species or environments. What is needed is a consistent, reliable, accurate measurement paradigm for captive animal welfare that can be used with a variety of species.

An animal's activity budget is considered to be one of the most important welfare measures (Young, 2003). If it is evident through behavioural observation that all of a captive animal's natural motivations, or ethological needs, are being fulfilled, then welfare can be said to be optimal. Often, captive animals forage and consume food for only a few minutes per day, whereas in the wild this would take up much of their time. Increasing foraging time can be a useful starting point in improving welfare as health problems in captive animals are often caused by an imbalance in nutritional intake versus exercise.

Stereotypic behaviour is another very commonly used welfare indicator for captive animals. If animals are not able to perform behaviours that they are strongly motivated to do, abnormal behaviours such as stereotypic behaviour can develop (Mason, 1991). Stereotypic behaviour is defined by Mason (1991) as unvarying, repetitive behaviour patterns with no obvious goal or function. Broom (1983) suggests that as a criterion, more than 10% of an animal's waking life spent in stereotypic behaviour is an indicator of poor welfare. Mason *et al.* (2007) call for 'zero tolerance' of stereotypic behaviour in captive animals, and argue that EE has had only limited success to date in eliminating stereotypic behaviour once it becomes established.

According to Dawkins (1988), physical health is one of the least controversial and easily measured welfare assays. This can be done in a variety of ways, including visual assessment, body weight and stress hormone analysis. Visual assessment is easier to carry out on large numbers of animals, such as those in farms or laboratories, or animals that are difficult to handle, but can miss the small changes that bioassays can reveal. However, Fox *et al.* (2006) note that measuring glucocorticoids as indicators of negative states is somewhat problematic as corticoids can indicate arousal as well as stress.

4.2.4 Cognitive Bias as a Welfare Measurement Tool in Non-Human Animals

A new development in the study of animal welfare has come from work with depressed and anxious humans. Humans experiencing a positive affective state tend to interpret ambiguous stimuli in a more positive way, i.e. they are

more optimistic, than people experiencing a negative affective state such as depression or anxiety (Eysenck *et al.*, 1991). This can be used to assess mental state in a non-verbal way. There is currently a great deal of work being conducted with animals using this paradigm. As Mendl *et al.* (2009) reason, if you can accurately assess this two-way flow between cognitive processes and emotional state in humans, it is possible that the same phenomenon exists in other species and can be assessed in a similar way. Boissy *et al.* (2007a) identify cognitive processes as having a key role in the perception of events and situations by animals, which in turn affects and is affected by their emotional state. They suggest that both cognition and emotion must be considered when assessing animal welfare.

Bateson and Matheson (2007) investigated the effects of environmental enrichment on the optimism of captive starlings, using a cognitive bias paradigm. They used a task based on taste aversion, training the birds using a white/dark grey, go/no go design, then presenting stimuli of ambiguous grey shading, to see if the starlings would interpret them as positive or negative stimuli, which enabled their classification as 'optimistic' or 'pessimistic'. The results of the study were that starlings exposed to enriched conditions, and then moved to unenriched conditions were significantly more pessimistic in their interpretation of ambiguous stimuli than starlings that experienced the conditions in the opposite order.

4.2.5 Task Complexity

Another issue that must be investigated is whether the time subjects spend interacting with enrichment is connected with welfare. If increased interaction is found to be linked with increased welfare, what factors are involved in increasing interaction? If subjects spend more time interacting with more complex enrichment, is this due to the simple fact that food is more difficult to extract, or is it that more complex puzzles stimulate their interest for longer?

4.2.6 Group and Individual Presentation

Additionally, as captive animals are housed in both group and individual settings, I wanted to see if group or individual presentation of enrichment has any bearing on its effectiveness in improving welfare. Pigeons have relatively

stable social hierarchies, with dominant individuals occupying the central (safest) roosting sites and group position. Generally there are two types of pigeon foragers – producers and scroungers (Giraldeau and Lefebvre, 1986). Young pigeons begin as scroungers, and some progress to producers, who are more skilled at finding food. Giraldeau and Lefebvre (1987) found that when a group of sixteen captive pigeons were given a novel foraging task involving pecking tubes to release food, 97% of all food discoveries were made by only two pigeons, the producers of the group, which the other pigeons, known as scroungers, preferentially followed around the enclosure. The scroungers failed to learn the method of food discovery until the producers were removed from the group, at which time another pigeon assumed the role of producer.

Dominance does not seem to be a deciding factor in determining producers and scroungers; Giraldeau and Lefebvre (1987) established that the pigeons had a stable hierarchy by observing the priority rank of access to freely available food over several weeks and found no link between dominance and food discovery role. Giraldeau and Lefebvre (1987) then investigated the learning ability of another group of pigeons by placing them in individual cages and using a pigeon ‘tutor’ in an adjacent cage to demonstrate the same tube pecking task. Half of the group were able to scrounge food from the demonstrator, while the other half were not. While 100% of the non-scrounge group learned to open the tube to get a food reward, only 25% learned the task in the scrounge group, suggesting that if scrounging is possible, pigeons generally prefer to do this rather than learn to find food for themselves. In experiment one, this thesis, subjects experienced competition upon opening a S+ cup, as other individual scroungers could feed from the cup once opened.

Crony *et al.* (2007) found that in chickens given a discrimination task to learn, dominance was not correlated with task acquisition. However, Baymann *et al.* (2007) found that in goats, lower-ranking individuals engaged less with a discrimination task and learned at a slower rate than higher-ranking individuals when their social group was mixed and relocated. The authors suggest that increased agonistic encounters brought about by the reorganisation of the hierarchy accounts for this deficit, and suggest that groups should remain stable in composition and housing if possible for CE to be beneficial.

Therefore, there are several elements that may interact in determining if group or individual presentation of enrichment is more effective in improving

welfare. If, in pigeons, dominance has little effect upon foraging success in a group, and pigeons prefer to scrounge if possible, it would be interesting to see if welfare was improved more in group- (naturally preferred by flocking animals, a safer mode of foraging) or individual- (likely to obtain more food) foraging task situations as this was not examined by Giraldeau and Lefebvre (1987).

4.2.7 Aims and Hypotheses

Experiment One

Group housing is a common way to house many captive animals, as it has been shown to improve the welfare of captive animals by creating a more natural living environment (e.g. Rault, 2012). Manteuffel *et al.* (2009) Kalb and Puppe (2010) and Zebunke *et al.* (2013) provided CE to group-housed individuals, all with positive effects on the welfare of captive farm animals. The current experiment therefore aimed to investigate whether providing captive, group-housed laboratory pigeons with CE would be beneficial in terms of welfare, thus investigating the effects of CE upon an avian, laboratory species, and to compare this with the effects of PE, and with baseline measures, in an ABACA repeated-measures design with individuals acting as their own controls, as this paradigm has not yet been attempted by any other researchers.

In an attempt to provide clear, reliable welfare measurement for the effects of EE, four measures of welfare were taken in both enrichment conditions, as well as in all three baseline conditions. These measures were activity budget, physical condition, stereotypic behaviour and 'optimism' as measured using a cognitive bias task based on Bateson and Matheson (2007). The enrichment items were as similar as possible, in terms of size, placement, time available, rewards, and ease of use, facilitating direct comparison.

The main predictions for this experiment were that during the CE condition of the experiment, subjects would show positive changes to their activity budget, would perform less stereotypic behaviour, and be more 'optimistic' (Bateson and Matheson, 2007) when presented with ambiguity, as compared to baseline conditions. It was also expected that these measures would show some improvement in the PE phase of the experiment, but that benefits would be shorter-lived as habituation was expected to occur more

rapidly. It was also expected that physical condition, stereotypic behaviour and optimism would be correlated, i.e. a subject in worse physical condition would perform more stereotypic behaviour and be less optimistic than a subject in better physical condition.

Experiment Two

A second experiment was designed using 'low-complexity' and 'high-complexity' CE devices, in order to find out whether it was solely more time spent extracting food from the device (in which case both low- and high-complexity CE devices would improve welfare the same amount), or more cognitive power used to interact with a device that brought about improvements in welfare (in which case the high-complexity CE device would be associated with a greater increase in welfare).

The aims of the study were to investigate the welfare benefits of two (low and high complexity) CE devices; and to ascertain whether the mode of presentation (group or individual) had any effect on engagement or welfare. Specifically, I hypothesized that both low and high complexity CE would be associated with a decrease in stereotypic behaviour and an increase in optimism compared with baseline measures, similar to the findings in experiment one. However, I expected that subjects in the high-complexity group would show greater welfare improvements, as higher complexity would lead to increased engagement and less habituation over time.

It was also predicted that presentation of enrichment (group versus individual) would have an effect on welfare. Either group enrichment would be more beneficial because of a feeling of safety in numbers and/or opportunities for social interaction and facilitation; or individual enrichment would create a safe environment without competition from scroungers or dominant individuals controlling access to the devices, which would allow for individual learning. Dominance and learning ability was also measured, and it was predicted that either lower-ranking pigeons would be slower at learning and engage less with the CE devices than higher ranking pigeons, consistent with findings in goats by Baymann *et al.* (2007); or that dominance, learning ability and engagement with the CE devices would not be correlated, similar to the findings of Giraldeau and Lefebvre (1987) in pigeons and Croney *et al.* (2007) in chickens. I therefore compared the welfare improvements found during experiment one with the

findings of experiment two in order to find out which mode of presentation conferred the greatest welfare improvements.

4.3 Method

4.3.1 Design (Expt. 1)

Experiment one took place between June-October 2009 and took 121 days to complete. The experiment had a repeated-measures design, with five conditions: baseline one, CE, baseline two, PE, and baseline three. The order was fixed for all subjects due to the fact that subjects were housed communally and the enrichment was presented within their home environment. The dependent variables measured were activity budget, stereotypic behaviour, physical condition, cognitive bias and use of enrichment. There was a seven day gap between each condition where the pigeons were left to settle with no enrichment or interference apart from the normal husbandry routine.

The experiment followed this timescale:

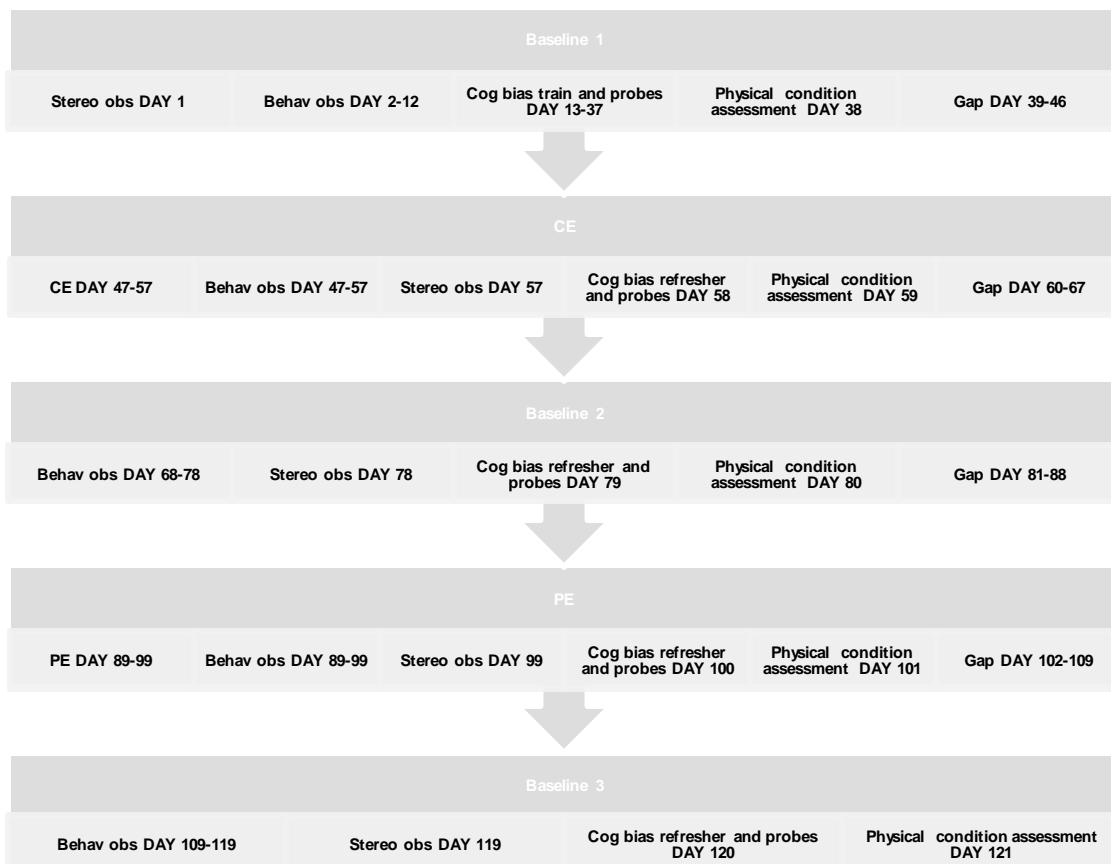


Fig. 1: Repeated-measures design of experiment one.

4.3.2 Design (Expt. 2)

Experiment two took place between May-July 2010 and took 36 days to complete. The experiment had a mixed design, with a baseline measurement phase, two experimental conditions (a) low-complexity cognitive enrichment device (LCCED) and (b) high-complexity cognitive enrichment device (HCCED) and a final baseline measurement phase. The dependent variables measured were stereotypic behaviour, cognitive bias, dominance, learning ability and use of enrichment. These measures were then compared with findings from experiment one to investigate the effects of group and individual presentation of enrichment.

The experiment followed this timescale:

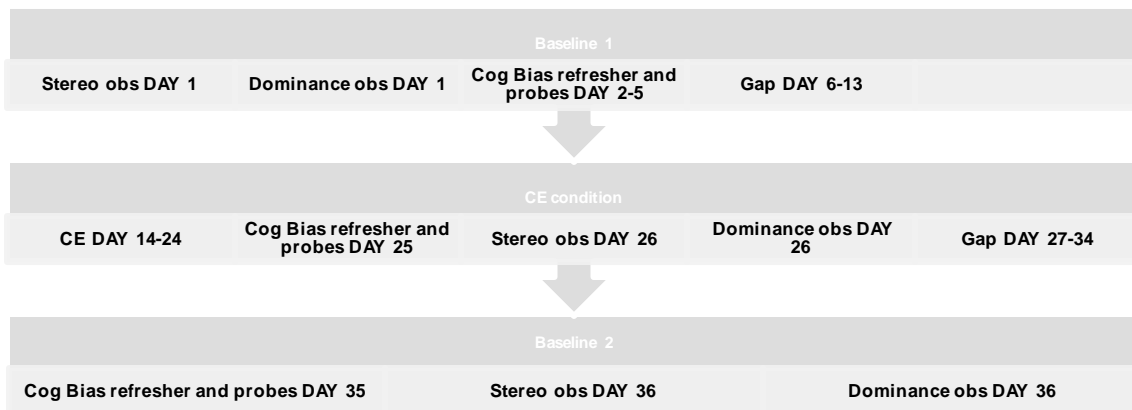


Fig. 2: Mixed repeated-measures design of experiment two.

4.3.3 Subjects and Housing (Expt. 1 and 2)

Sixteen pigeons in good health and normal weight range were studied altogether. They were procured in 2008 from a local pigeon fancier. The sex-ratio was approximately 1:1. Their enclosure was situated within the Washington Singer Laboratories at The University of Exeter in a windowless room (see Fig. 3), and was approximately 3m long by 1m wide, and 2.5m high. It was artificially lit on a 12:12 schedule, from 7am to 7pm including 15 minutes lightening and darkening to mimic dawn and dusk. It was separated from another aviary containing another 16 pigeons by a wood and chicken-wire frame. There were perches and shelves of various heights in the aviary. The pigeons had access to a maintenance ration of proprietary pigeon grain mix once daily at 4pm, scattered so they could forage for it - the floor was covered with horse bedding which provided an ideal substrate for pigeons to forage in. They also had free access to grit and water, and a water bath for bathing.

During the experiments, the pigeons were not involved in any other research, although they had all had experience with cognitive research in the past, ending two months prior to this set of experiments. Pigeons were identified using coloured rings and plumage variation. The same subjects were used in both experiments one and two, thus enabling the data from the two experiments to be compared.



Fig. 3: The aviary on the right in the picture held the 16 subjects, and was separated from the left aviary by a wire mesh. Visible at the front of the picture are the individual cages the subjects were trained to fly into for a food reward.

4.3.4 Reliability Assessment (Expt. 1 and 2)

All observations, training and other protocols during experiment one and two described here were conducted by one researcher (L. Millar). Inter-observer reliability was assessed by filming one hour of pigeon behaviour with no experimental interventions and asking a colleague (L. Halliday) to record behaviour on a focal-animal, instantaneous point sampling basis each minute, then comparing this with my own observations. Intra-observer reliability was assessed by observing the recording three weeks later to see if I recorded the same behaviours. Pearson's correlations were conducted using frequencies of behaviour for intra-observer ($r=0.997$, $p<0.001$) and inter-observer ($r=0.999$, $p<0.001$) reliability and found that although there were four instances of

disagreement in total there was a high degree of intra- and inter-observer reliability.

4.3.5 Activity Budget Measurement (Expt. 1)

During each of the five conditions, all sixteen pigeons were observed in their home environment using focal sampling, through a cardboard barrier placed at the door of the aviary to minimise visual disturbance. Behaviour observations took place in 10 minute blocks, using instantaneous point sampling on each minute between 7am and 7pm (the daylight hours for the pigeons) over ten days in order to minimise the effects of events occurring on a single day, for example cleaning out the aviaries and air-duct filter maintenance which tended to be quite disruptive. There were 72 10-minute observation blocks altogether. The pigeons were all given ID numbers, and were randomly selected for each 10 minute observation block using the research randomiser www.researchrandomizer.org. This meant that each pigeon was observed for ten minutes a mean of five times (range 4-6) across daylight hours, in each of the five experimental conditions. During the two enrichment conditions (CE and PE), observations took place at times when the enrichment devices were not in the enclosure. Behaviours measured were inactivity, locomotion, autogrooming, alertness, foraging, social, vocal and agonistic. An activity budget was constructed for each pigeon for each experimental condition and proportions of time spent engaged in each activity were calculated.

4.3.6 Physical Condition Measurement (Expt. 1)

The physical condition score sheet used in this experiment was designed for ease of use by laboratory technicians to assess the physical health of the population of laboratory pigeons each week, and as such involved only simple categorical observations (designed by L. Millar as part of laboratory technician duties).

It had the following categories: eyes (bright, clear?), beak (clean, undamaged?), feet (claws undamaged, toes uncalloused, straight and mobile?) wings/legs (strong, straight, using all limbs?), feathers/skin (undamaged, full quota, any evidence of feather-pulling?), parasite load (feather mites evident?), excreta (normal colour/texture/quantity?), eating/drinking (observe eating and drinking, normal speed and vigour of consumption, swallowing normal?),

muscle tone (strong, fine muscle tone? Palpate leg and wing, test pull-back) and weight rating (under/normal/over weight compared to own free-feeding weight, is there a covering of fat over breast-bone?). The scores for each section were between one and three, with one meaning excellent, two meaning minor problems and three meaning severely compromised requiring assistance. Each subject was thoroughly checked and an overall physical condition score was then calculated for each pigeon. This assessment was repeated once per experimental condition, at the end of each phase.

4.3.7 Stereotypic Behaviour Measurement (Expt. 1 and 2)

Due to stereotypic behaviour not being performed by all subjects, the behaviour occurring at relatively low frequencies compared with more common behaviour such as resting, and incidences of it occurring at specific times of the day rather than spread out (less than 3% of all stereotypic behaviour seen was performed at other times of day), all occurrences of stereotypic behaviour made by any subject were recorded between the hours of 2pm-7pm on a single day in each condition for both experiments (see design section for more information) through a cardboard barrier placed at the door of the aviary to minimise visual disturbance. This time of day was when subjects tended to engage in stereotypic behaviour. ID numbers, start/stop times and type of stereotypic behaviour was noted. Stereotypic behaviour was defined operationally for this group of pigeons as spot-pecking (repeated pecking lasting for more than ten seconds at one spot in the enclosure such as the dividing wall, a shelf or perch), and sham-drinking (pecking repeatedly at the water with a closed beak and no swallowing motion seen). Other stereotypic behaviour such as air-pecking were seen rarely in this group of pigeons and were recorded as 'other'. The proportion of time spent engaged in stereotypic behaviour between the hours of 2pm and 7pm was then calculated for each subject in each experimental condition. Video records were made of each type of stereotypic behaviour observed, these are available electronically.

4.3.8 Cognitive Bias Task Training and Probe Trials (Expt. 1 and 2)

Apparatus

Stimuli for the cognitive bias task (shown in Fig. 4) were made of thick card and protected with clear plastic. They were discs 80mm in diameter, and there were six in total. The 100% shaded (black) and 0% shaded (white) stimuli were the training stimuli, where half the subjects were given a small food reward – five hemp seeds - for removing the black stimulus from the apparatus, and half the subjects were given the same reward for removing the white stimulus. Subjects were allocated the same stimulus as S+ in both experiments. For the probe trials, there were four stimuli, one 20% shaded grey, one 40% shaded grey, one 60% shaded grey and one 80% shaded grey (see Fig. 4). The apparatus was a wooden block with a small feeding dish attached, where food rewards were placed directly (see Fig. 5). Experimenter effects were minimised by the experimenter moving back behind a cardboard barrier and videoing each trial. Subjects had to flip the shaded disc off the dish in order to access the hidden food rewards.

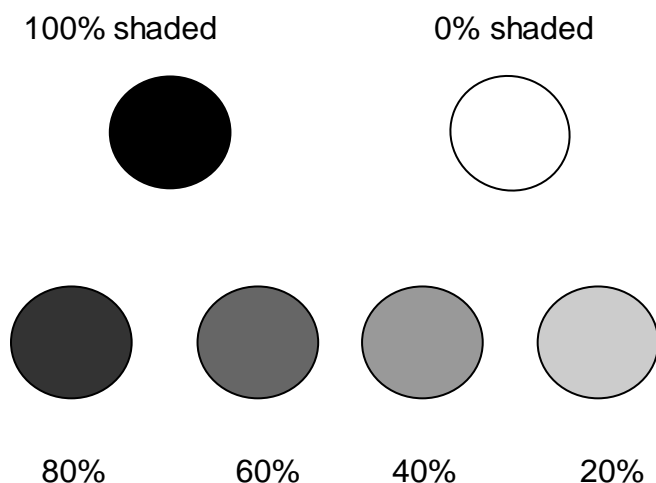


Fig. 4: Stimuli used in the cognitive bias task.

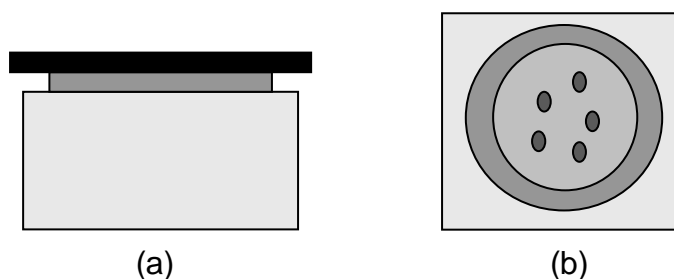


Fig. 5: (a) side view and (b) top view of apparatus – a wooden block with metal feeding dish attached, stimulus placed on top, (b) shows food reward visible when stimulus is flipped.

Procedure

After the first baseline welfare observations were carried out, training for the cognitive bias task began. All the pigeons had already been trained to fly to individual cages for a small food reward. As much as possible, the pigeons were not handled, and instead flew in and out of the individual cages themselves. This was to keep stress to a minimum. The decision by the pigeon to fly into a cage was deemed to be evidence of volunteering to participate. If a subject did not engage with the task within 10 minutes, or showed signs of distress (flapping, pacing along the sides of the cage or panting) they were released to fly back to the aviary. Apart from the time spent in a cage training for the task or being given enrichment, they continued to live communally as normal. During the training phase, the order in which the pigeons were trained

was randomised each day. After training, subjects were released from the testing cages and they flew back to their home aviary.

Training

The cognitive bias task was based on the paradigm set by Bateson and Matheson (2007), with the following alterations. The task used white and black stimuli as S+ and S-, counterbalanced between the subjects. Each subject was given one training session per day consisting of eight trials, with one stimulus presented at a time. S+ was always rewarded with five hemp seeds. S- was never rewarded, and in addition a 30 second time-out was imposed if subjects flipped a S-. Subjects were given 20 seconds to either flip the S+ and consume the reward, or to ignore the S-. The choice made by the subject was noted, and the apparatus removed and replaced with the next stimulus and reward if any. The inter-trial-interval was 40 seconds. The training phase continued until the pigeons scored 7/8 correct choices in three consecutive sessions, upon which they were moved to the probe phase. Before the next probe session could take place, subjects were given a refresher training session until they reached the same 7/8 correct in three consecutive sessions. The same applied at the beginning of experiment two. A record was made of the number of sessions needed for each pigeon to reach criterion initially – this was used as a measure of their learning ability. The decision to use time-out instead of a more noxious punishment such as electric shock was an ethical one, as Lea (1979) points out that it is our duty as researchers to implement methodology that causes the least amount of suffering in the subject, especially as these experiments were aiming to improve welfare.

Probes

After reaching criterion in the training phase, subjects were given a 'probe session' where they were shown eight S+, eight S- and four each of the four grey stimuli, in a randomised order, totalling thirty-two trials. The reason for this 'supersession' is that conducting four separate probe sessions of eight trials over four days would not capture a 'snapshot' of the subject's affect directly after an experimental manipulation.

Ambiguous stimuli did not have a time penalty imposed if flipped and were rewarded on a variable interval schedule of one reinforced to three

unreinforced during a session of thirty-two trials, so that subjects experienced all ambiguous stimuli as reinforced on a 0.25 basis. This was to prevent the subjects from learning to interpret all ambiguous stimuli as either S+ or S-. The proportion of ambiguous stimuli flipped by subjects was taken to be their 'optimism' score. This protocol was repeated once during each condition of the experiments, with subjects being given probe sessions in the morning following the end of the enrichment conditions, in order to prevent confounds from either the enrichment being currently available or subjects starting to be affected by its absence.

4.3.9 Dominance Observations (Expt. 2)

Dominance interactions were recorded on the first day of baseline one, the last day of the CE phase and the last day of baseline two using all-occurrence sampling through a cardboard barrier placed at the door of the aviary to minimise visual disturbance on one day for three hours after scattering food on the floor of the enclosure at 4pm until darkness at 7pm. Specific interactions looked for were supplanting, displacement, pecking and fights, and each interaction usually had a clear winner. These interactions were used to construct an interaction matrix, and then the pigeons were ranked according to the number of interactions they had won and lost. As some subjects did not interact during the observation time, the formula by David (1987) was used:

$$DS = W + W_2 - I - I_2$$

Where W indicates wins by the individual, W_2 indicates the wins by their opponents, I indicates losses by the individual and I_2 indicates losses by their opponents. I used this formula because it was found by Gammel *et al.* (2003) and De Vries *et al.* (2006) to be one of the most effective hierarchy formulas for analysing tied or unknown relationships. The procedure for calculating David's Score (DS) for each subject in a group was carried out by weighting each individual's relative success by the success of their opponents. This means that a win against a higher ranking opponent is valued higher than a win against a lower ranking opponent and a hierarchy can then be constructed including all individuals (see David, 1987; and De Vries *et al.*, 2006, for more details on how to calculate David's Score). The dominance hierarchy data were then used to assess how dominance interacted with welfare and the provision of CE.

4.3.10 Cognitive Enrichment (CE) Condition (Expt. 1)

Apparatus

The layout of the cognitive enrichment device (CED) was based upon an experiment carried out by Wills *et al.* (2009) which made use of an array design set out on a touchscreen. The CED (shown in Fig. 6) was constructed from a wood and cork notice board (a), with ten plastic cups painted either orange or blue, cut down to half size with a small cup attached underneath (b). The orange cups were each food-rewarded with approximately ten highly preferred seeds (hemp). The cups snapped back into position after being opened which meant that the rewards were not visible until the cups were opened, they were difficult to access, and took longer to empty than if the cups did not close.

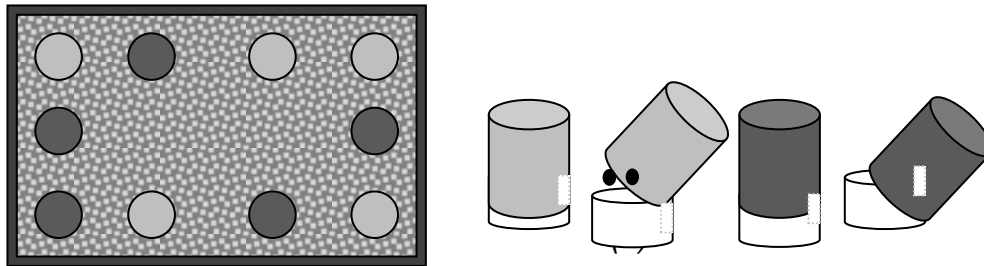


Fig. 6: (a) Top view of the CED, constructed from a cork notice-board with cups attached using split pins, light grey represents orange (S+) and dark grey represents blue (S-). (b) Side view of the open and closed cups with food reward visible in the orange cup.

Procedure

For this experimental condition, the pigeons continued to live communally in their home aviary, and for ten days, between 1.30-2.30pm (staggered in order to prevent schedule stereotypies developing and to allow for behaviour observation when the device was unavailable), the CED shown in Fig. 5 was put in to the aviary for 60 minutes. The device and surrounding area was filmed through a cardboard barrier placed at the door of the aviary to minimise visual disturbance, and a one-zero sampling method was used to record whether during each minute, each of the subjects interacted with the device or not. The device was placed in the same spot in the middle of the

aviary floor each day, and rotated 90° each day to avoid positional biases developing.

Twenty percent of the birds' daily ration was placed in the orange cups, and the CED was placed in the middle of the floor of the enclosure each day. On day six, the pigeons succeeded in finishing all the food rewards after twenty minutes, so from day seven onwards the total food reward was split into three and the orange cups were filled with a reduced reward at zero, twenty and forty minutes. This was to ensure that the pigeons did not lose interest in the CED due to there being no reward. The CED was removed from the enclosure after an hour, then the pigeons were scatter-fed the remaining 80% of their ration at 4pm.

4.3.11 Physical Enrichment (PE) Condition (Expt. 1)

Apparatus

The layout of the physical enrichment device (PED) was based upon the design of the CED, to keep the two enrichments as similar as possible in all ways except that there were no specific cues associated with the food rewards, therefore there was no opportunity for discrimination. The PED was constructed from the same size cork notice board as the CED (see Fig. 7). A lattice of cardboard strips forming a loosely-structured grid was stapled on to the wooden frame for the birds to forage through to get at the hemp seeds below.

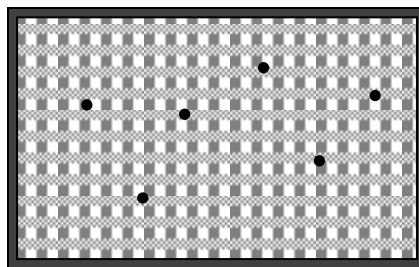


Fig. 7: The PED with scattered hemp seeds amongst strips of cardboard which were stapled on to the cork board.

Procedure

For this experimental condition, the pigeons continued to live communally in their home aviary, and for ten days, between 1.30-2.30pm

(staggered in order to prevent schedule stereotypies developing and to allow for behaviour observation when the device was unavailable), the PED shown in Fig. 7 was put in to the aviary for sixty minutes, the device and surrounding area was filmed through a cardboard barrier placed at the door of the aviary to minimise visual disturbance, and a one-zero sampling method was used to record whether during each minute, each of the subjects interacted with the device or not. The device was placed in the same spot in the middle of the aviary floor each day, and rotated 90° each day to avoid positional biases developing.

Twenty percent of the pigeons' daily food ration was scattered into the cardboard grid. From the seventh day, the ration was divided into three portions, then scattered onto the cardboard grid at zero, 20 and 40 minutes of the enrichment hour each day to mirror the CE condition. The PED was removed from the enclosure after an hour, and the pigeons were then scattered the remaining 80% of their ration at 4pm.

4.3.12 Cognitive Enrichment (CE) Condition (Expt. 2)

Apparatus

Two CE devices were designed for this experiment based on the CE device used in experiment one: (a) a Low-Complexity Cognitive Enrichment Device (LCCED) and (b) a High-Complexity Cognitive Enrichment Device (HCCED), see Fig. 8. The enrichment devices were each made with a cork notice board, to which 12 painted plastic cups were pinned, allowing one side to flip up when pecked to reveal a smaller open cup inside which could hold a small food reward. In the LCCED the reward was placed under six random cups each day, in the HCCED the rewards were placed under the light or dark green cups depending on the subject's random allocation to 'light green or dark green positive', and the cups were moved around the board in the same linear configuration but in different orders, with no more than three of the same colour cups in a row.

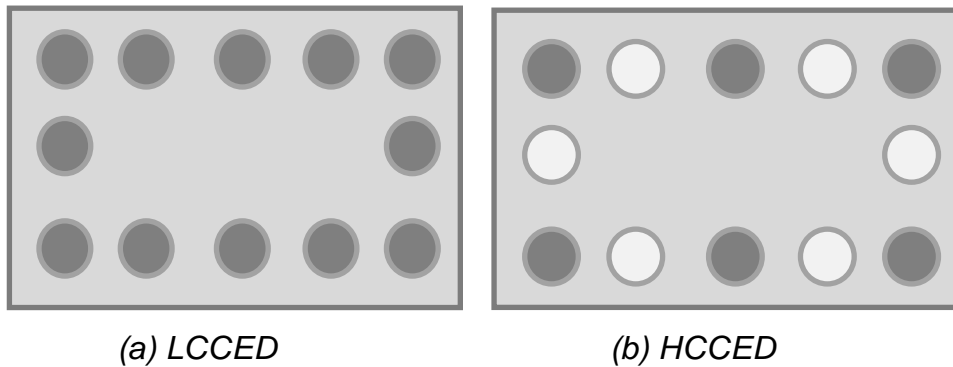


Fig. 8: (a) LCCED with 12 dark green plastic cups, and (b) HCCED with six light green and six dark green plastic cups, attached using split pins, under which food rewards were hidden.

Procedure

For the enrichment phase of the study, the pigeons continued to live communally in their home aviary, and were allocated randomly into two groups (N=8 in each) which determined which enrichment device they would be presented with during the CE condition, either the LCCED or the HCCED.

All subjects were already trained to fly to the individual cages for a small food reward. As much as possible, the pigeons were not handled, and instead flew in and out of the cages themselves. This was to keep stress to a minimum. The decision by the pigeon to fly into a cage was deemed to be evidence of volunteering to participate. If a subject showed signs of distress (flapping, pacing along the sides of the cage or panting) they were released to fly back to the aviary. Otherwise, each subject was given ten minutes per day for ten days in an individual cage (a medium size dog crate) with one of the enrichment devices depending on their allocation, their behaviour was filmed, experimenter effects were minimised by the experimenter moving back behind a cardboard barrier during each enrichment session, and every peck they made upon the enrichment device was recorded.

4.4 Results

All data were analysed using SPSS v.19. When possible, parametric statistical tests were used, however some datasets showed significant departures from normality despite arcsine transformations, therefore non-parametric tests were implemented where necessary. Where multiple comparisons were made,

Bonferroni significance corrections were used. Mdn signifies median, R is range, M is mean and SE is standard error.

4.4.1 Activity Budget (Expt. 1)

As can be seen from Fig. 9, there were significant differences in subjects' activity budgets due to experimental manipulations.

Activity budget was analysed using Friedman tests for all five phases of the experiment. There was an effect of experimental condition upon the behaviour of the subjects, specifically upon autogrooming ($X^2(4) = 13.3$, $p=0.001$); alertness ($X^2(4) = 11.90$, $p=0.003$) and agonistic behaviour ($X^2(4) = 21.64$, $p < 0.001$) (see Fig. 7), bonferroni corrected significance set at $p = 0.006$. There were no significant differences between baseline and experimental conditions for inactivity, foraging, affiliative, vocal display or stereotypic behaviour (separate from the all-occurrence observations of stereotypic behaviour discussed below).

The three baseline conditions for autogrooming, alertness and agonistic behaviour were then analysed using Friedman tests to see if subjects returned to baseline levels of each behaviour after each enrichment condition. There were no significant differences between the three baselines (autogrooming baseline one Mdn 0.19, baseline two Mdn 0.18, baseline three Mdn 0.17, $X^2(2) = 1.37$, $p = 0.51$; alertness baseline one Mdn 0.07, baseline two Mdn 0.08, baseline three Mdn 0.07, $X^2(2) = 1.18$, $p = 0.56$; agonistic baseline one Mdn 0.01, baseline two Mdn 0.02, baseline three Mdn 0.01, $X^2(2) = 0.62$, $p = 0.73$) so in order to aid the analysis of the effects of the two enrichment conditions, I combined the three baselines to give an 'overall baseline' proportion for each behaviour for each subject (See Fig. 10).

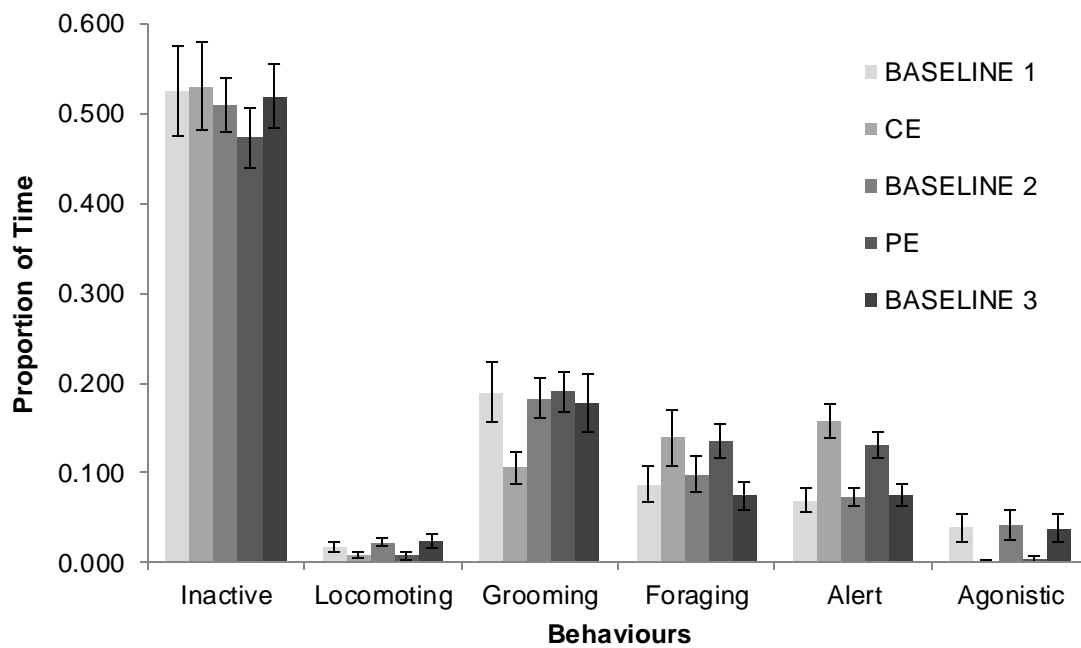


Fig. 9: The five experimental conditions in order, showing changes in proportion of time spent engaging in each main activity during each experimental condition. Error bars represent standard error.

Affiliative, vocal display and stereotypic behaviour are not included in Fig. 9 or 10 because they were performed at very low levels compared with other categories of behaviour. Fig. 10 reveals significant differences between baseline and experimental conditions in autogrooming, alertness and agonistic behaviour. Wilcoxon tests confirmed this, these are described below.

A Wilcoxon test revealed that subjects groomed themselves significantly less in the CE condition (Mdn 0.10, R 0.00-0.21) than in the baseline conditions (Mdn 0.20, R 0.05-0.43, $Z = -3.05$, $p = 0.002$) and the PE condition (Mdn 0.19, R 0.01-0.39, $Z = -2.50$, $p = 0.012$). There were no significant differences between the PE condition and baseline conditions ($Z = -0.09$, $p = 0.93$).

A second Wilcoxon test revealed that subjects were significantly more alert in the CE condition (Mdn 0.12, R 0.06-0.33, $Z = -2.54$, $p = 0.01$) and the PE condition (Mdn 0.12, R 0.04-0.27, $Z = -2.90$, $p = 0.004$) than in the baseline conditions (Mdn 0.06, R 0.03-0.14). No significant differences were found between the PE and CE conditions ($Z = -1.40$, $p = 0.16$).

A third Wilcoxon test revealed that subjects engaged in significantly less agonistic behaviour in the CE condition (Mdn 0.00, R 0.00-0.01, $Z = -2.53$, $p = 0.01$), and PE condition (Mdn 0.00, R 0.00-0.06, $Z = -2.37$, $p = 0.02$) than in the

baseline conditions (Mdn 0.02, R 0.00-0.19). No significant differences were found between the CE and PE conditions ($Z = -1.07, p = 0.29$).

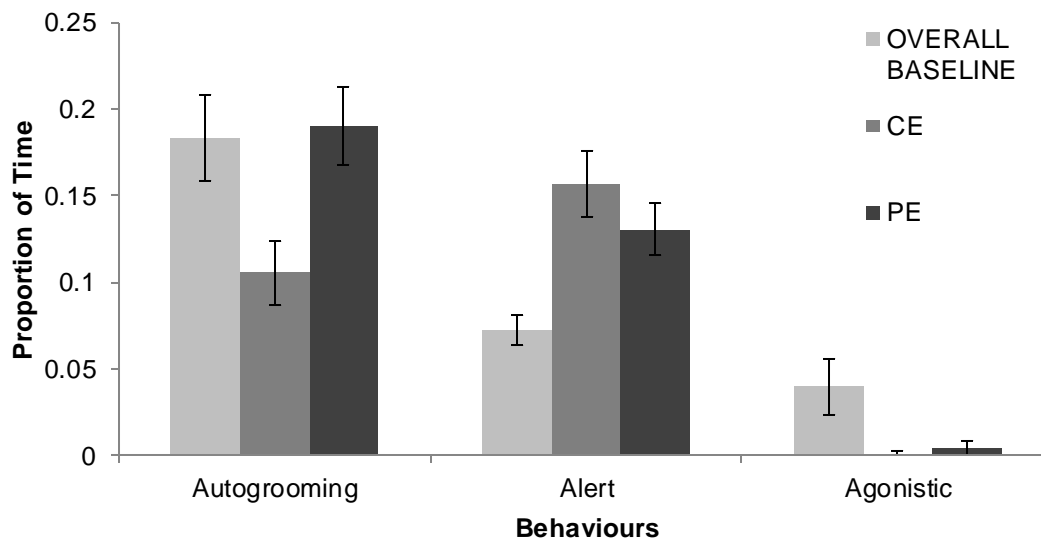


Fig. 10: Significant behaviour changes during the experimental conditions. The three baseline conditions have been combined to produce an overall baseline as there were no significant changes in behaviour over the three baseline measurements. Error bars represent standard error.

4.4.2 Physical condition (Expt. 1)

The physical condition of subjects did not differ significantly across experimental conditions. The physical condition scores were analysed using a repeated measures ANOVA, but there were no significant differences found in the physical condition of subjects across conditions ($F(3, 21)=2.27, p>0.05$).

Fig. 11 shows the relationship between mean physical condition and mean stereotypic behaviour across all conditions. Between subjects, physical condition was found to correlate with stereotypic behaviour: subjects in worse physical condition tended to perform more stereotypic behaviour ($r=0.57, p=0.01$).

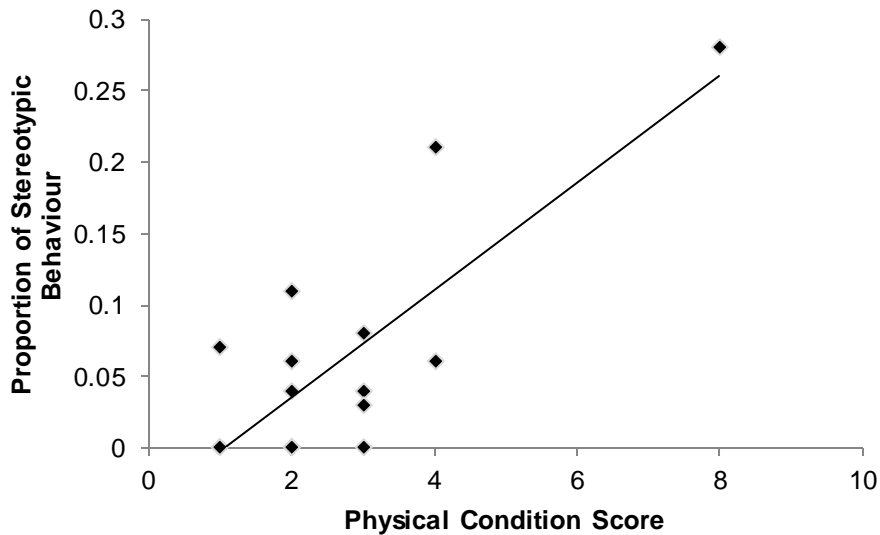


Fig.11: Overall physical condition scores and the overall proportion of stereotypic behaviour seen in each subject.

4.4.3 Stereotypic behaviour (Expt. 1)

Stereotypic behaviour was observed in ten of the sixteen subjects in baseline one, in those ten subjects performance ranged from 3-28% of total time observed spent in stereotypic behaviour (M = 10%). In the CE condition, stereotypic behaviour was observed in nine subjects, R = 0-11%, M = 3%. In the PE phase, stereotypic behaviour was again observed in nine subjects, R = 0-38% M = 11%. See Fig. 12 for more information.

There were significant departures from normality so the data from only those subjects that showed any sign of stereotypic behaviour in the first observation were used, N=10. This allowed the assumption of normality to be realised. The assumption of sphericity was violated ($X^2(9) = 30.40, p = 0.001$) therefore the Huynh-Feldt correction was applied.

A repeated measures ANOVA indicated that there was a significant main effect of experimental condition upon the proportion of time subjects were seen engaging in stereotypic behaviour in the all-occurrence sessions dedicated to measuring stereotypic behaviour in each condition ($F(1.77, 15.91) = 5.72, p = 0.02$).

No significant differences were found between the three baseline measures of stereotypic behaviour (baseline one M=0.10, SE=0.03; baseline two M=0.10, SE=0.02; baseline three M=0.10, SE=0.03; sphericity assumed

$F(2,18) = 6.95, p = 0.51$). The baseline conditions were therefore combined to create a mean stereotypic behaviour proportion for baseline conditions.

A repeated measures ANOVA was then conducted on the proportion of stereotypic behaviour for the overall baseline condition ($M=0.10, SE=0.03$), the CE condition ($M=0.03, SE=0.01$) and the PE condition ($M=0.11, SE=0.04$): Mauchly's test indicated that the assumption of sphericity had not been violated ($X^2(2)=5.37, p=0.07$). There was again a significant main effect of condition on the proportion of stereotypic behaviour observed ($F(2,18)=6.22, p=0.009$). A contrast revealed a significant difference between the overall baseline condition and the CE condition ($F(1,9)=16.03, p=0.003$). This means that subjects were found to perform significantly less stereotypic behaviour in the CE condition than in the baseline conditions.

A second contrast revealed another significant difference between the CE condition and PE condition ($F(1,9)=5.76, p=0.04$). This means that subjects were found to perform significantly less stereotypic behaviour in the CE condition than in the PE condition.

A third contrast revealed no significant difference between the proportion of stereotypic behaviour observed in baseline conditions and the PE condition ($F(1,9)=0.02, p = 0.88$).

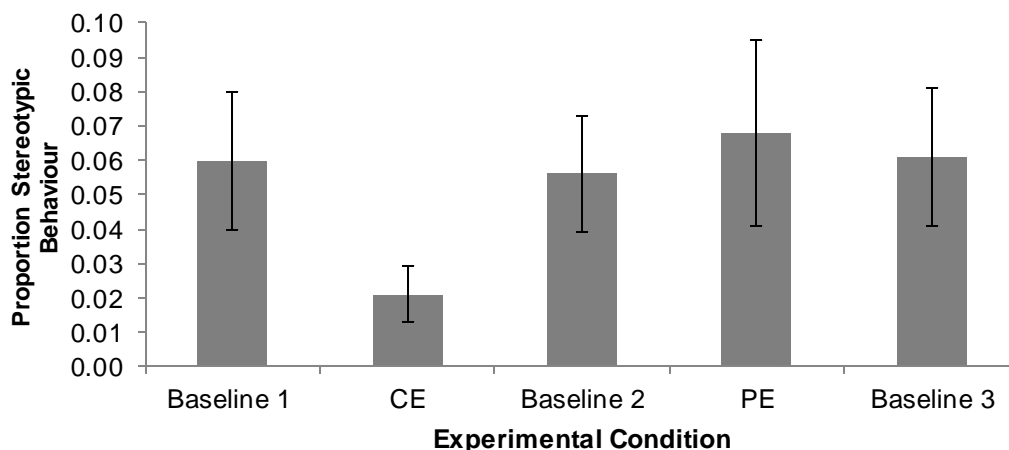


Fig. 12: Mean proportion of stereotypic behaviour observed in all subjects in each condition. Error bars represent standard error.

4.4.4 Optimism (Expt. 1)

Three individuals did not learn the cognitive bias task and were therefore excluded from the analysis. The data showed significant departures from normality despite arcsine transformation therefore non-parametric tests were used.

Fig. 13 indicates that there were significant differences in the proportion of ambiguous grey probe stimuli chosen by subjects, hereafter known as their optimism scores. Subjects ranged from 0-100% in their optimism scores, $M = 59\%$ in baseline conditions, 80% in the CE condition and 67% in the PE condition.

A Friedman test indicated that there was a significant effect of experimental condition upon subjects' optimism scores: $X^2(4) = 3.09$, $p = 0.011$.

There were no significant differences found between the three baseline optimism scores (baseline one Mdn 0.69, R 0.00-1.00; baseline two Mdn 0.69, R 0.00-1.00; baseline three Mdn 0.83, R 0.00-1.00; $X^2(2) = 4.19$, $p = 0.12$). The baseline scores were therefore combined to create a mean optimism score for each subject during baseline conditions.

A Friedman test revealed that there was a significant difference between the combined baseline optimism score and the CE and PE conditions ($X^2(2) = 7.24$, $p = 0.021$). Wilcoxon tests showed that the difference lay between the baseline optimism score and the CE condition (All Baseline Mdn 0.71, range 0.00-1.00, $Z = -2.67$, $p = 0.008$). This means that subjects had significantly higher optimism scores in the CE condition than in the baseline conditions.

A second Wilcoxon test did not reveal any difference between optimism scores in the baseline conditions and PE condition ($Z = -0.98$, $p = 0.33$).

A third Wilcoxon test also revealed no significant difference between optimism scores in the CE condition and the PE condition ($Z = -1.26$, $p = 0.21$).

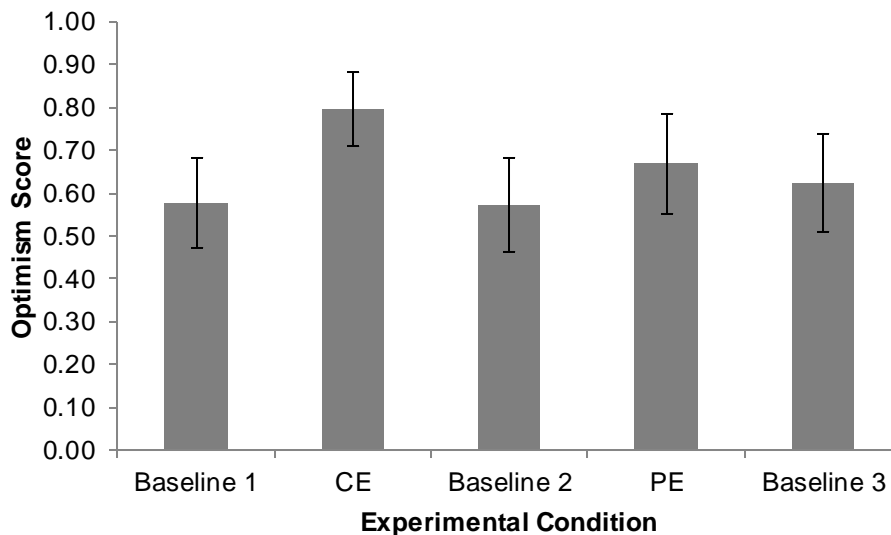


Fig. 13: Mean optimism scores in each experimental condition. Error bars represent standard error.

The optimism scores were then broken down into probe gradients from 20% positive to 80% positive. Note that the experiment was counterbalanced so that half the subjects received 0% shaded as positive whilst the other half received 100% shaded as positive – the colour palette in Fig. 14 represents all subjects' scores from closest to negative to closest to positive over the five experimental conditions. A repeated measures ANOVA was carried out on the individual probes chosen in each experimental condition. There was no interaction between experimental condition and probe choice ($F(12,180) = 0.64$, $p = 0.81$) meaning that the type of probe choices did not change significantly over the course of the experiment. This indicates that the reward schedule was sufficient to keep all four probe choices ambiguous over repeated presentations. If, for example, the proportion of 20% positive choices had decreased steadily over the course of the experiment, it would be evident that subjects were learning that the probe was associated with a lower chance of reward than other probes. Examination of Fig. 14 suggests that the probe closest to positive was consistently chosen by subjects at the highest rate, while the two probes closest to negative were chosen least often in all experimental conditions. A between-groups ANOVA shows that subjects' choice of the four probes differed significantly ($F(3,192) = 13.80$, $p < 0.0001$). Table 1, below, compares the probes with t-tests, to see where the differences lie. Significant differences were found between all of the probes except the two least positive

and the two most positive probes, i.e. subjects interpreted those as functionally similar.

	20	40	60	80
20(M=2.17,SE=0.24)	0	NS	t(64) = 4.15, p < 0.001	t(64) = 5.12, p < 0.0001
40(M=2.32,SE=0.24)		0	t(64) = 3.01, p = 0.004	t(64) = 3.74, p < 0.0004
60(M=2.72,SE=0.22)			0	NS
80(M=2.91,SE=0.22)				0

Table 1: The four probe types with t-tests, bonferroni correction set at 0.01.

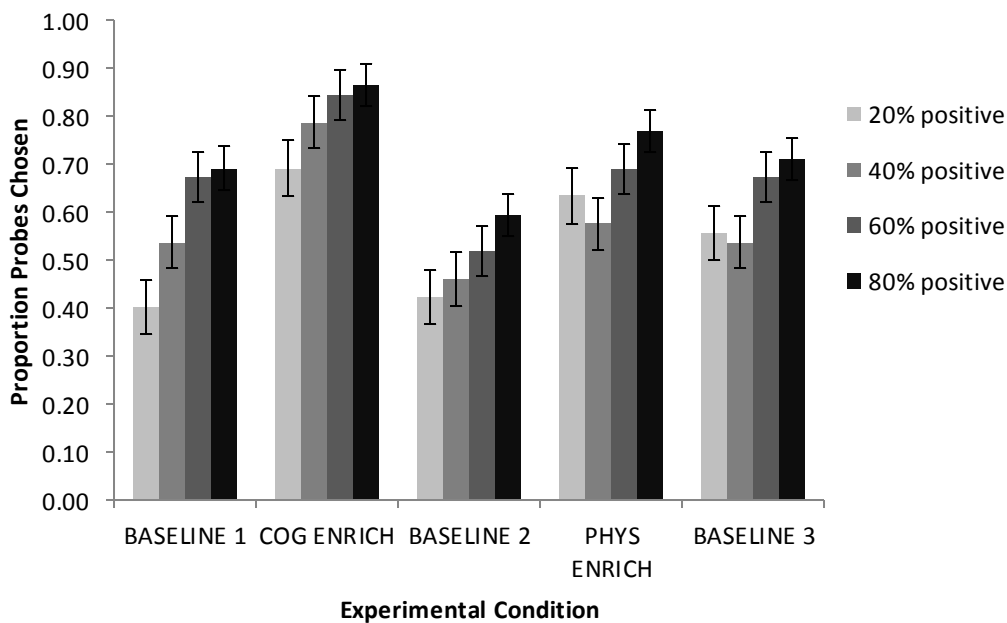


Fig. 14: Proportion of each probe gradient chosen in each experimental condition. Error bars represent standard error.

Optimism was not correlated with the performance of stereotypic behaviour at all (for mean optimism and stereotypic behaviour across all conditions: $r=0.20$, $p=0.45$), meaning that individuals that performed higher levels of stereotypic behaviour were not more likely than individuals that performed lower levels of stereotypic behaviour to have lower optimism scores, either in baseline or enrichment conditions.

4.4.5 Use of Enrichment (Expt. 1)

All subjects interacted with both the CED and the PED. Fig. 13 indicates that subjects used the CED significantly more than the PED overall, measured using a one-zero recording method during each of the 60 minutes per day the devices were available. There was a significant effect of condition on enrichment use: the CED (Mdn 0.51, R 0.33-0.84) was used by the subjects significantly more than the PED (Mdn 0.24, R 0.18-0.58; $Z = -3.47$, $p = 0.001$).

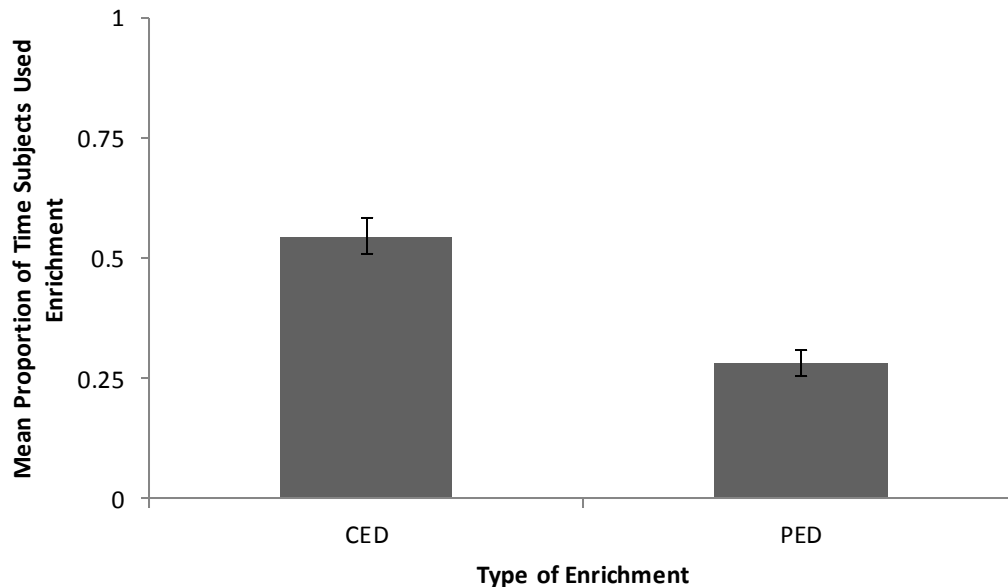


Fig. 15: Mean proportion of use of enrichment in each enrichment condition overall, error bars represent standard error.

Fig. 15 suggests that subjects showed more interest in the CED than the PED from around day four, although this difference reduces towards day ten. There was a significant effect of day on enrichment use ($F(9,270)=17.57$, $p<0.001$, Huynh-Feldt correction for sphericity), meaning that, taken together with fig. 16, as the days progressed, enrichment use increased. There was also a significant interaction between day and condition ($F(9,270)=12.90$, $p<0.001$, Huynh-Feldt correction for sphericity), suggesting that over time the use of the CED particularly increased.

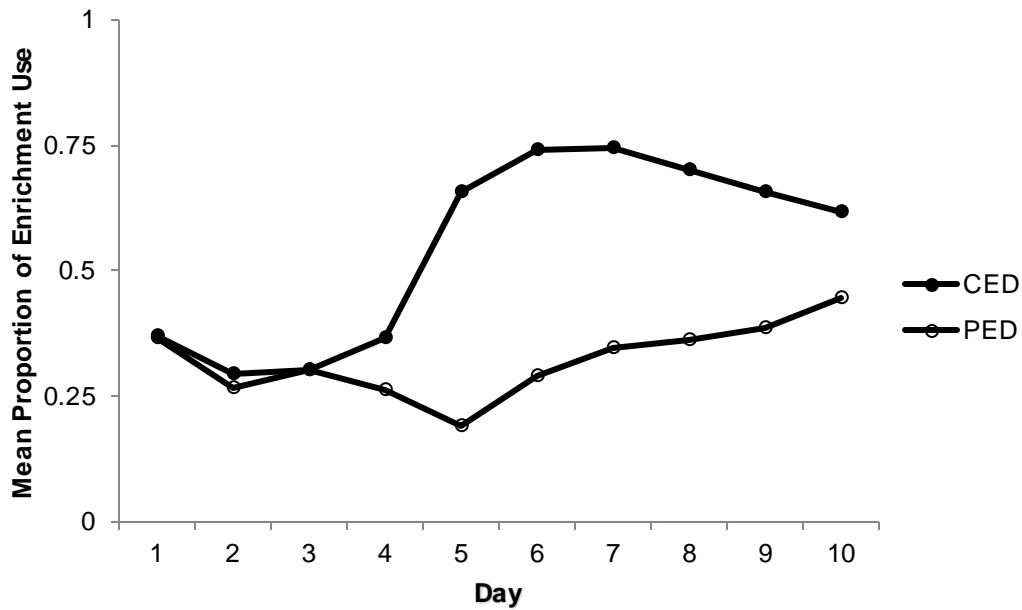


Fig. 16: Mean proportion of use of enrichment in each enrichment condition over time.

4.4.6 Predicting Welfare Improvement (Expt. 1)

Subjects were classified as low users (less than the mean of 55% for the CED, or less than the mean of 28% for the PED) or high users (mean of 55% or more for the CED, or 28% or more for the PED) of each enrichment device. Use of enrichment did not appear to affect optimism or stereotypic behaviour significantly.

However, a path analysis (shown in Fig. 17) was carried out using enrichment use and enrichment type as predictors of optimism, and indicated that optimism was affected mainly by engagement time rather than by the nature of the device itself. Stereotypic behaviour could not be predicted accurately by either factor.

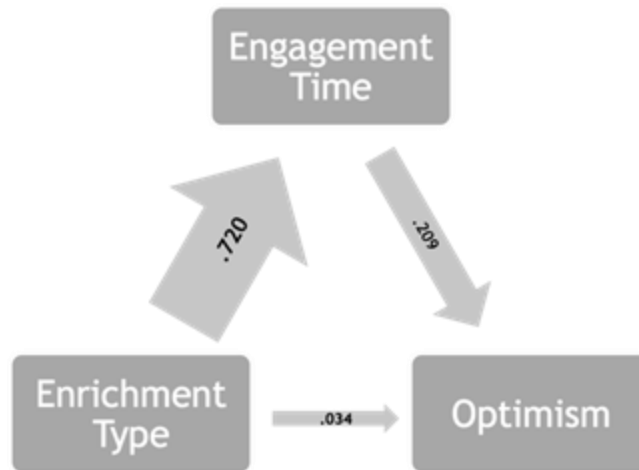


Fig. 17: Path analysis of the use of the CED and PED.

4.4.7 Stereotypic behaviour (Expt. 2)

Of the ten subjects observed to engage in stereotypic behaviour, performance ranged from 2-72% of total time observed in baseline one (M 30%). In the CE phase, stereotypic behaviour ranged from 0-24%, M 8%, and it was evident in only six subjects. In baseline two, stereotypic behaviour observed ranged from 0-40% in eight subjects, M 18%. Fig. 18 shows the amount of stereotypic behaviour performed by subjects in each condition.

There were significant departures from normality so the data from only those subjects that showed any sign of stereotypic behaviour in any condition were used, N=10 (the same individuals as in experiment one, this thesis). This allowed the assumption of normality to be realised. The assumption of sphericity was unviolated ($X^2(2) = 5.79, p = 0.06$).

A repeated measures ANOVA with group (low-complexity versus high-complexity CE devices) as a between-groups factor indicated that there was a significant main effect of experimental condition on the level of stereotypic behaviour performed by subjects, $F(2,18) = 6.11, p = 0.009$.

Contrasts revealed that the differences lay between baseline one (M = 0.30, SE = 0.09) and the CE condition (M = 0.08, SE = 0.03; $F(1,9) = 7.30, p = 0.02$, and between the CE condition and baseline two (M = 0.18, SE = 0.05; $F(1,9) = 5.51, P = 0.04$). There were no significant differences between baseline one and baseline two ($F(1,9) = 4.19, p = 0.07$). This means that stereotypic behaviour was significantly reduced in the CE condition, and rose again in baseline two. However, the type of enrichment, either high or low complexity,

had no significant effect on stereotypic behaviour ($F(2,8)=2.75$, $p=0.09$). This means that although the subjects benefitted significantly from access to CE, the level of complexity did not affect stereotypic behaviour.

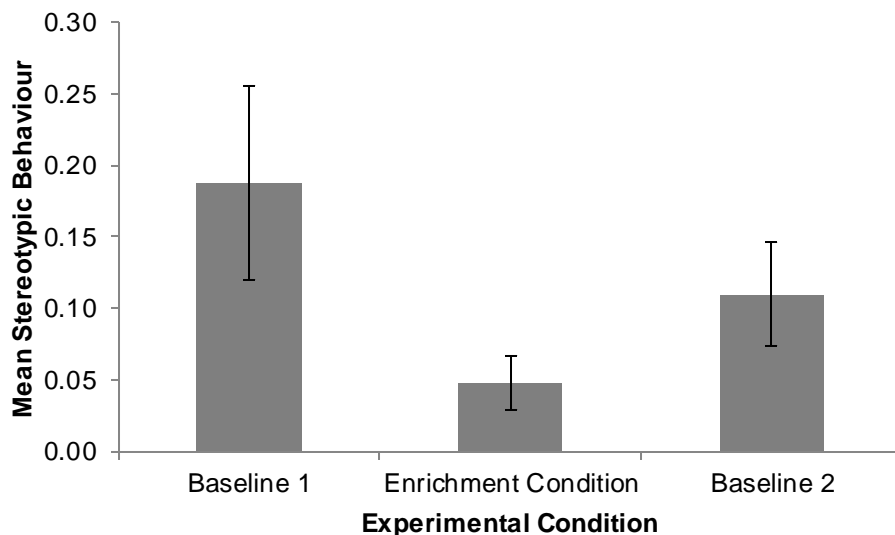


Fig. 18: Proportion of stereotypic behaviour observed in all subjects in the baseline (unenriched) conditions and in the CE condition (combined low and high complexity conditions). Error bars represent standard error.

4.4.8 Optimism (Expt. 2)

Three subjects did not learn the discrimination in experiment one, this thesis, therefore optimism was measured in the remaining thirteen subjects. Fig. 19 shows the mean optimism scores for subjects in each condition.

A repeated measures ANOVA indicated that there was a significant effect of CE on the level of optimism shown by the subjects in interpreting ambiguous probes (baseline one optimism $M=0.56$, $SE=0.11$; CE optimism $M=0.79$, $SE=0.12$; baseline two optimism $M=0.64$, $SE=0.11$ $F(2,22)=8.62$, $p=0.002$). This means that the level of optimism changed significantly between experimental conditions. Contrasts reveal that the significant differences lay between baseline one and the CE condition ($F(1,11)=15.95$, $p=0.002$); and between the CE condition and baseline two ($F(1,11)=6.95$, $p=0.02$). A final contrast revealed that optimism scores in baseline one and baseline two did not differ significantly ($F(1,11)=1.60$, $p=0.23$).

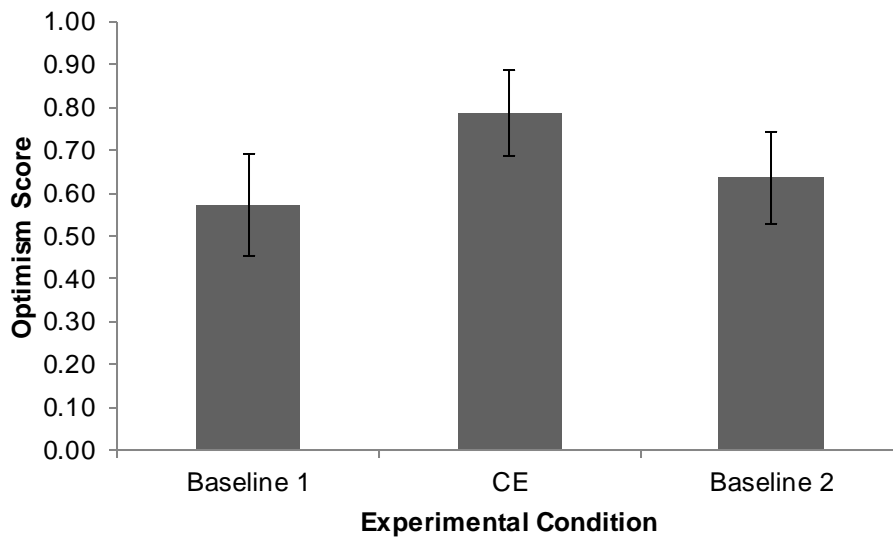


Fig. 19: Mean optimism scores for subjects in each experimental condition. Error bars represent standard error.

The optimism scores were broken down into probe gradients from 20% positive to 80% positive. Note that the experiment was counterbalanced so that half the subjects received 0% shaded as positive whilst the other half received 100% shaded as positive – the colour palette in Fig. 20 represents all subjects' scores from closest to negative to closest to positive over the three experimental conditions. A repeated measures ANOVA was carried out on the individual probes chosen in each experimental condition. There was no interaction between experimental condition and probe choice ($F(6,108) = 1.66$, $p = 0.14$) meaning that the type of probe choices did not change significantly over the course of the experiment. This indicates that the reward schedule was sufficient to keep all four probe choices ambiguous over repeated presentations. If, for example, the proportion of 20% positive choices had decreased steadily over the course of the experiment, it would be evident that subjects were learning that the probe was associated with a lower chance of reward than other probes. Examination of Fig. 20 suggests that the probe closest to positive was consistently chosen by subjects at the highest rate in all experimental conditions. A between-groups ANOVA shows that subjects' choice of the four probes did not differ significantly ($F(3,152) = 1.30$, $p = 0.28$). This means subjects interpreted all probes as functionally similar.

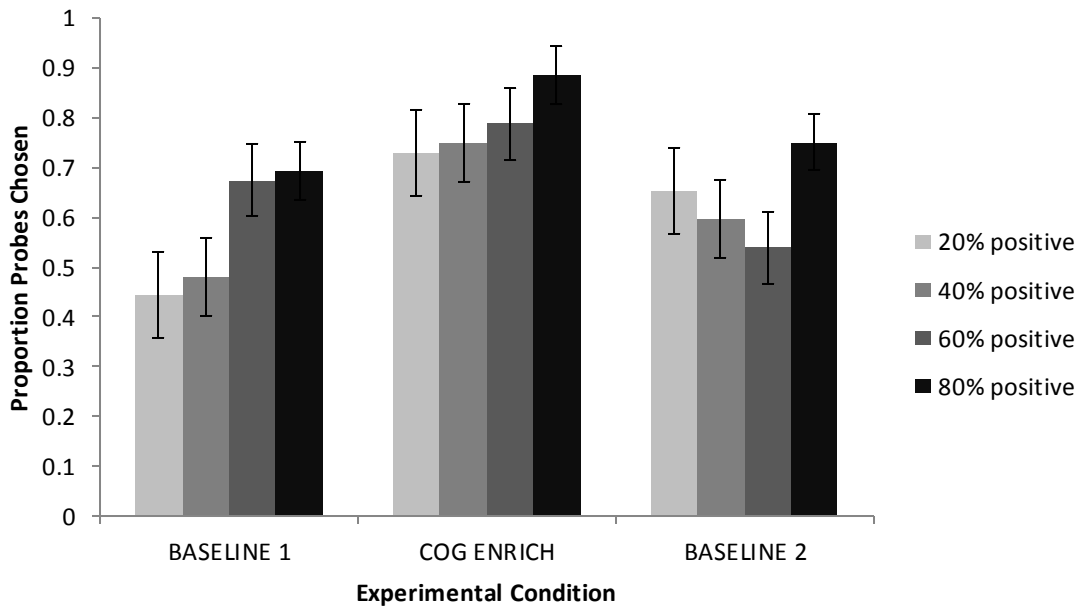


Fig. 20: Proportion of each probe gradient chosen in each experimental condition. Error bars represent standard error.

For reasons unknown, despite subjects being assigned to the high and low complexity groups randomly using researchrandomiser.org, the low-complexity group were notably (but not significantly: $F(1,12)=1.32$, $p=0.28$) less optimistic in the first baseline condition than the high-complexity group. There was a significant interaction between optimism and enrichment group (baseline one optimism low complexity $M=0.43$, $SE=0.17$; baseline one optimism high complexity $M=0.70$, $SE=0.16$; CE optimism low complexity $M=0.80$, $SE=0.16$; enrichment optimism high complexity $M=0.79$, $SE=0.14$; baseline two low complexity $M=0.48$, $SE=0.15$, baseline two high complexity $M=0.77$, $SE=0.14$; $F(2,22)=4.28$, $p=0.03$). This means that optimism varied significantly according to whether subjects experienced the low or high complexity enrichment, with a larger effect in the low-complexity CE group. Fig. 21 shows this relationship between CE complexity and optimism. Both enrichment types raised subjects' optimism scores to the same level, a rise of 36% in the low-complexity condition, and a rise of 10% in the high-complexity condition, to a high point of around 80% for both conditions.

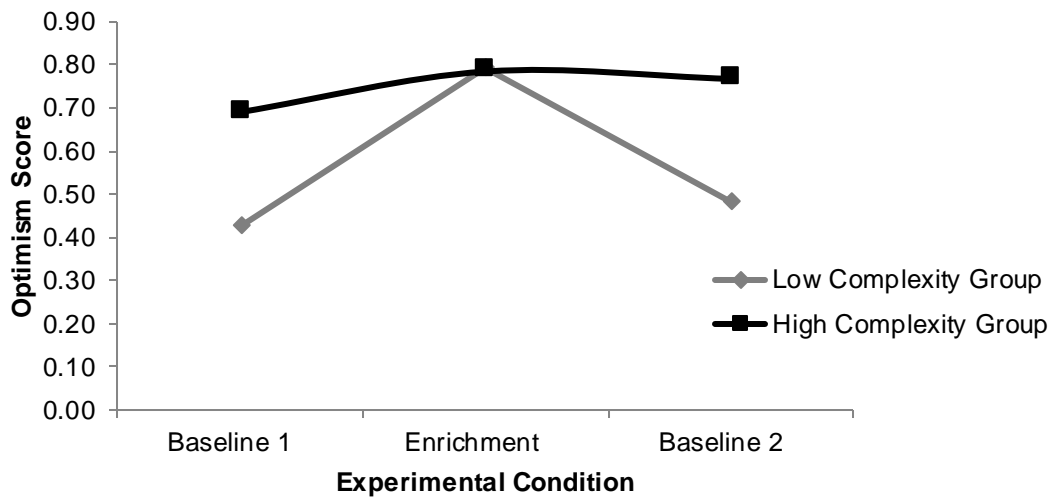


Fig. 21: Mean optimism scores for subjects in the low- and high-complexity enrichment groups over the baseline and experimental conditions.

4.4.9 Use of Enrichment (Expt. 2)

Use of enrichment was measured using pecks. Each peck by a subject on the enrichment device was recorded using a hand-held event logger. A one-way ANOVA revealed that there was no significant difference overall in the amount subjects used the CE in the low- and high-complexity groups (pecks to low-complexity CE device $M=1136$, $SE=1220$; pecks to high-complexity CE device $M=2522$, $SE=750$; $F(1,15)=2.57$, $p=0.13$).

Subjects were then categorised as low or high enrichment users (less or more than the median of 1369 pecks per session). Inspection of Fig. 22 shows that high use of enrichment was linked with increased optimism. A one-way ANOVA was carried out on the on the optimism scores of subjects categorised as low ($N= 8$) and high users ($N=8$) which indicated that the interaction between use of enrichment and optimism scores was significant ($F(2,22)=3.27$, $p=0.05$).

There were no links found between use of enrichment and stereotypic behaviour, learning ability or dominance.

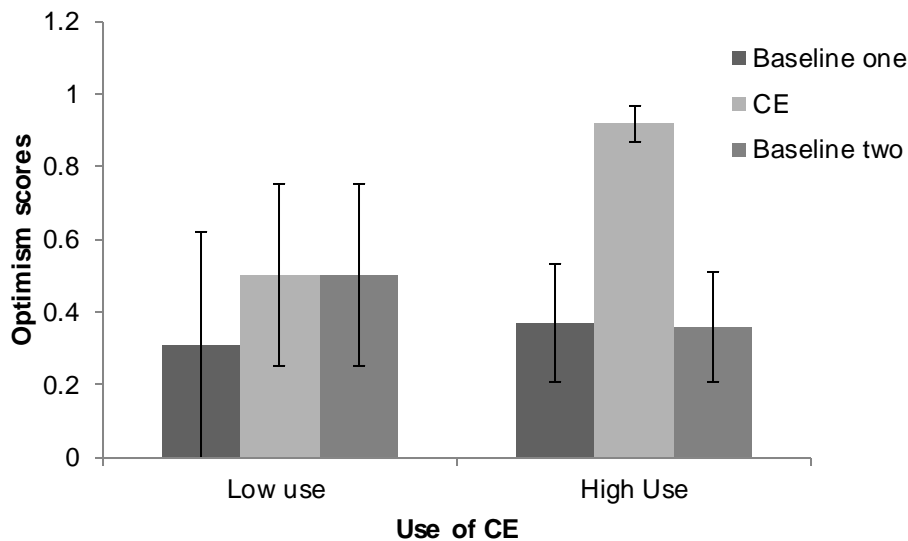


Fig. 22: Use of enrichment and optimism in subjects. Subjects categorised as high users showed a larger increase in their optimism scores during the CE condition than low users. Error bars represent standard error.

There were both low and high users in both enrichment groups (low-complexity/low-use N=5; low-complexity/high-use N=3; high-complexity/low-use N=4; high-complexity/high-use N=4). Over the ten days of enrichment, those in the low-complexity group ranged from a mean of 60-3475 pecks to the device, while those in the high-complexity group ranged from a mean of 41-5348 pecks to the device. However, examination of Fig. 22 indicates that overall, the high-complexity enrichment group used their enrichment to a greater extent over time than the low-complexity group. Subjects did not begin to properly interact with the enrichment devices until day 5, therefore analyses were restricted to days 5-10. Using a repeated measures ANOVA there was a significant main effect of day of enrichment presentation. The assumption of sphericity was violated, $X^2(14)=31.49$, $p=0.005$, therefore the Hyuhn-Feldt correction for sphericity was used ($F(2.86,40.05)=9.89$, $p<0.001$). This result, in conjunction with visual interpretation of fig. 23, suggests that, over time, both groups used their CE devices significantly more. There was also a significant interaction of day and group ($F(2.86,40.05)=3.19$, $p=0.04$; Hyuhn-Feldt correction for sphericity). This suggests that subjects in the high-complexity CE group interacted with their enrichment device significantly more between days 5-10, and their use of the enrichment increased steadily over time with no evidence of

habituation, compared with subjects' interaction with the device in the low-complexity CE group. Fig. 23 indicates that the low-complexity group seemed to begin to habituate to the CE device on the last two days of the experimental condition.

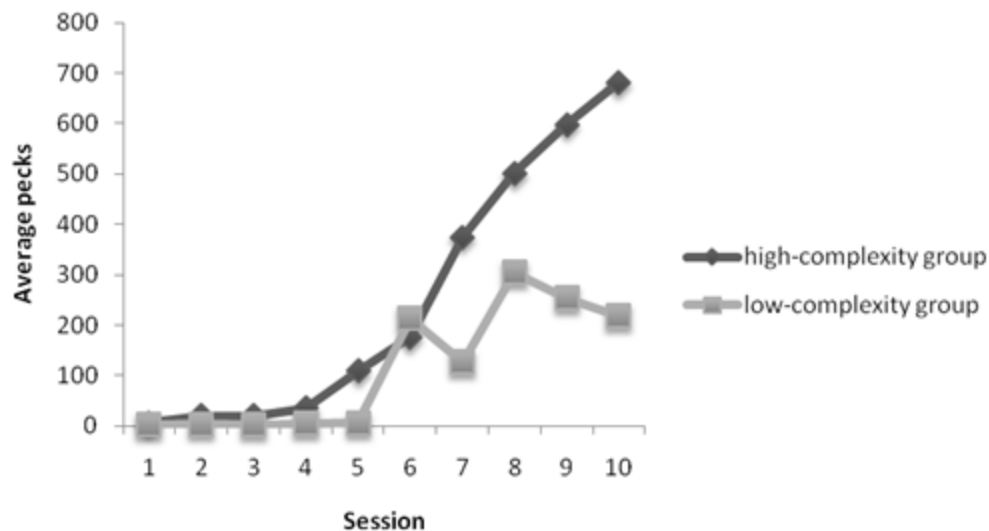


Fig. 23: Mean pecks upon the two types of enrichment, indicating use of enrichment, over the ten sessions.

4.4.10 Group versus Individual Enrichment Presentation (Expt. 2)

In order to compare group- and individually-presented CE, a repeated measures ANOVA was carried out on stereotypic behaviour and optimism scores using baseline one and CE condition data for group-presented CE (from experiment one) and individually-presented CE (from experiment two) for those individuals that performed stereotypic behaviour (N = 10).

There was a significant main effect of mode of presentation on stereotypic behaviour (experiment one baseline stereotypic behaviour $M=0.10$, $SE=0.03$; experiment one CE condition stereotypic behaviour $M=0.03$, $SE=0.01$; experiment two baseline stereotypic behaviour $M=0.30$, $SE=0.09$; experiment two CE condition stereotypic behaviour $M=0.08$, $SE=0.03$; the assumption of sphericity was violated $X^2(5)= 22.94$, $p<0.001$ therefore Huynh-Feldt correction for sphericity was used: $F(1.31, 9.14)=6.77$, $p=0.02$).

A contrast revealed that there was no significant difference between the baseline measures of stereotypic behaviour in experiment one and experiment two ($F(1,7)=5.66$, $p = 0.05$).

A second contrast revealed that there was no significant difference between the CE condition stereotypic behaviour in experiment one and experiment two ($F(1,7)=4.15$, $p=0.08$). This means that both group- and individually-presented CE reduced the incidence of stereotypic behaviour to similar levels for both experiments. Fig. 24 shows this reduction of stereotypic behaviour in both experiments.

There was a significant main effect of mode of presentation on optimism (experiment one baseline optimism $M=0.57$ $SE=0.11$; experiment one CE condition optimism $M=0.80$, $SE=0.09$; experiment two baseline optimism $M=0.57$, $SE=0.12$; experiment two CE condition optimism $M=0.79$, $SE=0.10$; $F(3,36)=7.47$, $p=0.001$).

A contrast revealed that there was no difference between the baseline measures of optimism in experiment one and experiment two (experiment one baseline optimism $M=0.57$, $SE=0.11$; experiment two baseline optimism $M=0.57$, $SE=0.12$; $F(1,12)=0.001$, $p=0.97$).

A second contrast revealed that there was also no difference between the CE condition optimism in experiment one and experiment two (experiment one CE optimism $M=0.80$, $SE=0.09$; experiment two CE optimism $M=0.79$, $SE=0.10$; $F(1,12)=0.05$, $p=0.83$). This means that optimism was a relatively stable measure in both experiments, with similar baseline measures that rose to similar levels during both group- and individually-presented CE. Fig. 24 shows the increase in optimism in both experiments.

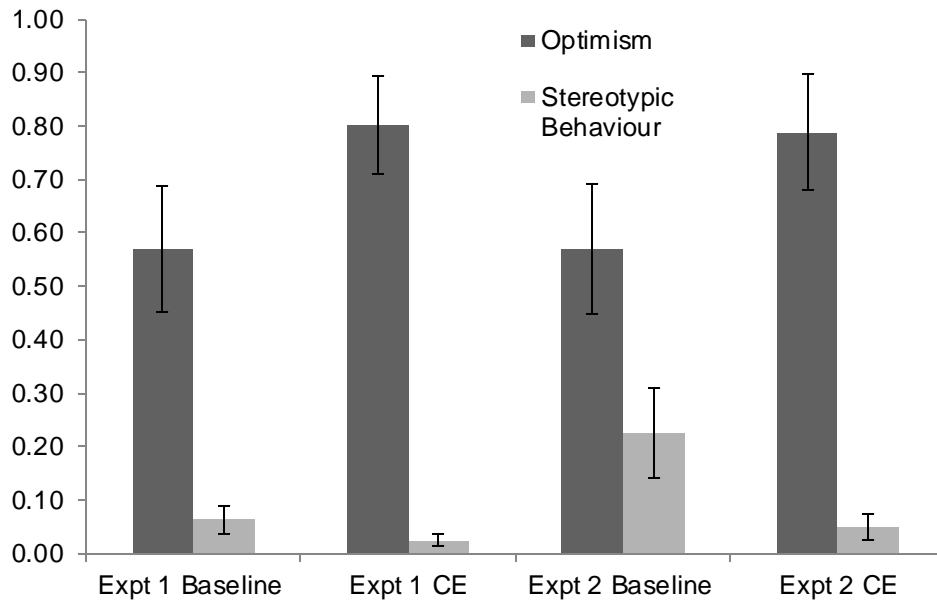


Fig. 24: Changes in stereotypic behaviour and optimism when presented with CE in a group (experiment one) and individually (experiment two). Error bars represent standard error.

4.4.11 Dominance (Expt. 2)

Subjects were found to have a strong linear hierarchy, with most agonistic encounters taking place with and between individuals ten, twelve and one. Interactions not involving these three individuals were rare. Table 2 shows the ranks of individuals in each experimental condition. From table 2 we can see that the hierarchy appeared stable over time, with individuals twelve, ten and one consistently occupying the top three positions, while individuals nine, four, eight and five occupied the middle ranks, and thirteen, two, fifteen, fourteen and seven were in the low ranks. There were two social climbers, individuals three and sixteen went from low- to middle-ranking over the course of the study, while individual six dropped from top to middle and individual eleven dropped from middle- to low-ranking.

	Mean Rank	Baseline Rank	CE Rank	Baseline 2 Rank
High	12	12	12	12
	10	6	10	10
	1	10	1	1
Medium	6	1	3	9
	9	4	9	16
	4	11	11	8
	11	5	6	6
	8	9	8	3
	3	8	4	4
	5	7	16	13
16	2	5	5	
Low	13	3	13	14
	2	13	15	2
	15	16	2	7
	14	14	14	11
	7	15	7	15

Table 2: Subjects' dominance hierarchy ranked by David's Scores for each experimental condition.

Fig. 25 shows the overall mean DS for each subject, while Fig. 26 shows the mean DS for all subjects in each condition. A repeated measures ANOVA indicated that subjects' DS varied significantly between experimental conditions ($F(2,30)=14.52$, $p<0.001$). Contrasts revealed that the difference lay between the CE condition and the baseline conditions: (baseline one mean DS=10.78, SE=2.05; CE mean DS=18.15, SE=2.63; baseline two mean DS=10.45, SE=2.71; contrast one between baseline one baseline two $F(1,15)=0.03$, $p=0.87$; contrast two between baseline one and CE condition $F(1,15)=22.638$, $p<0.001$; contrast three between CE condition and baseline two $F(1,15)=7.602$, $p=0.02$). This means that during the CE condition, individuals had significantly higher DS scores than in baseline conditions, which means there were more agonistic interactions overall during the CE period.

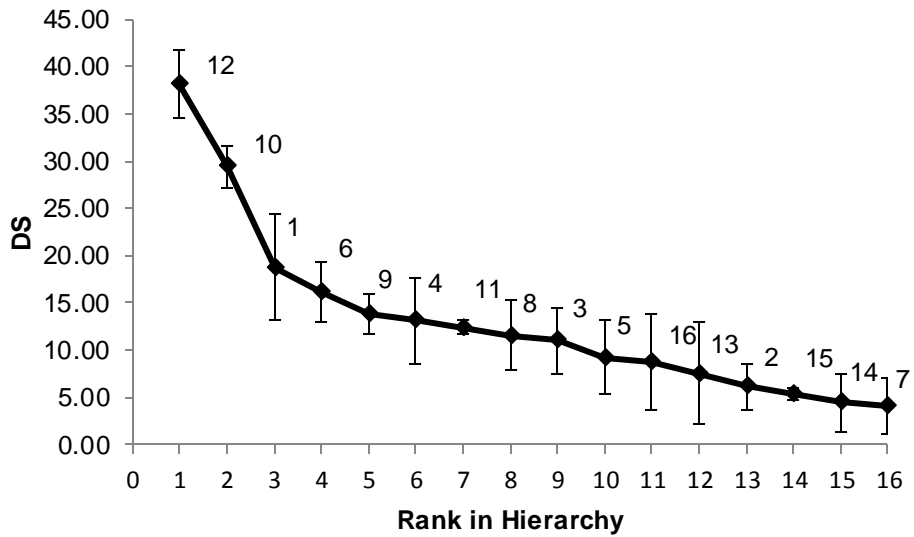


Fig.25: Dominance hierarchy constructed using mean David's Scores (DS) (David, 1987). Individual subjects shown as numbers from twelve (highest ranking) to fourteen (lowest ranking). Individuals twelve, ten and one (all mated males) were the highest ranking individuals with the remaining individuals occupying middle and subordinate positions. Error bars represent standard error.

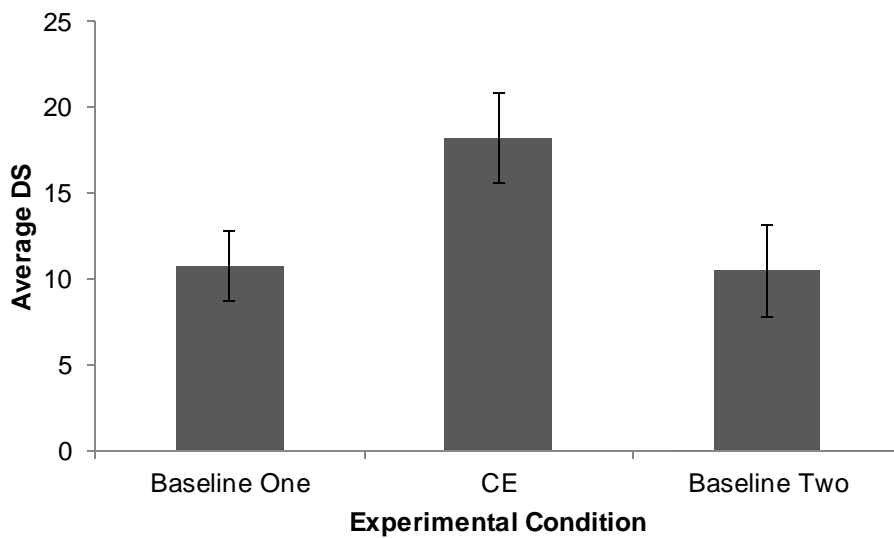


Fig. 26: Mean DS for each condition. DS was significantly higher in the CE condition which means that there were more agonistic encounters during that time.

Dominance measured during the CE condition was found to correlate with learning ability (see Fig. 27), with individuals occupying more dominant positions taking longer to learn the cognitive bias task to criterion in experiment

one, this thesis ($r=0.58$, $p=0.04$). Dominance did not correlate with optimism, stereotypic behaviour or use of enrichment in either group- or individually-presented enrichment, or low- and high-complexity enrichment.

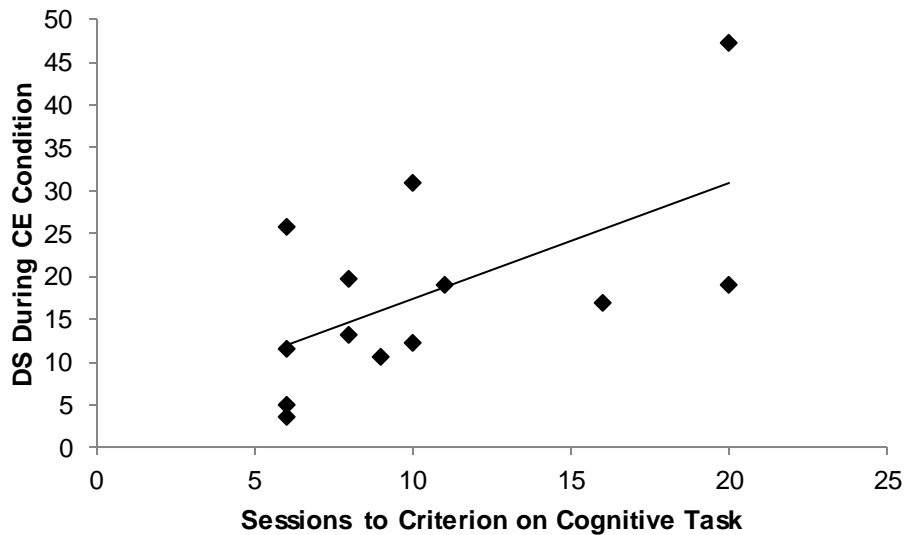


Fig. 27: Correlation between the DS for each individual during the CE condition and the sessions to criterion on the cognitive bias task. A higher number of sessions to criterion indicates a lower learning ability.

4.4.12 Predicting Welfare Improvement (Expt. 2)

A path analysis as shown in Fig. 28 was carried out using low- or high-complexity CE grouping, use of enrichment, learning ability and dominance data as predictors for subjects' change in optimism. Results indicated that optimism changes could only be predicted by use of enrichment, which was similar to findings in experiment one. Examination of fig.8 above indicates that high users of CE had higher optimism than low users during the CE condition. Stereotypic behaviour could not be predicted accurately, also similar to the finding in experiment one.

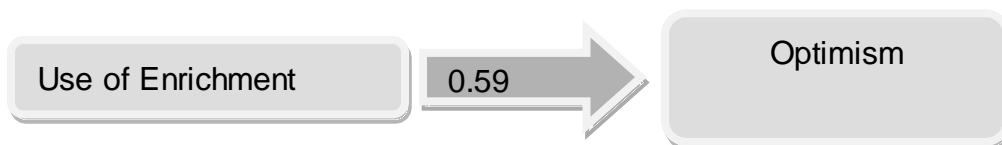


Fig.28: Path analysis of use of enrichment and corresponding improvement in optimism.

4.5 Discussion

In accordance with my main hypotheses, in experiment one I found that there were significant behavioural changes including reduced stereotypic behaviour, as well as an increase in subjects' optimism, during the CE condition when compared to the baseline conditions and PE condition.

I also found in experiment two that there were significant changes in the welfare of subjects, evident by reduced stereotypic behaviour, as well as an increase in subjects' optimism, in both low- and high-complexity CE conditions when compared to the baseline conditions. Subjects in the high-complexity enrichment group used their CE device significantly more than subjects in the low-complexity enrichment group, and higher use of both devices predicted higher optimism. Group and individual presentation of enrichment both had positive effects on the welfare of subjects.

4.5.1 Activity Budget (Expt. 1)

The activity budget of the pigeons changed significantly dependent on experimental condition. Significantly lower incidences of autogrooming were seen in the CE condition compared with the baseline or PE conditions. This reduction of self-directed behaviour suggests that the CE had a positive effect upon the subjects in terms of reducing chronic stress, which is consistent with Beerda *et al.* (1999) who found that dogs that experienced social, enriched housing performed less autogrooming than when in barren, isolated housing, and linked autogrooming with chronic stress.

Agonistic behaviour was also significantly reduced in both the CE and PE conditions compared with the baseline conditions, suggesting that both CE and PE are useful in reducing agonistic conflict in group-housed pigeons. This is consistent with Zebunke *et al.* (2013) who found that pigs provided with CE were involved in significantly fewer agonistic encounters than a control group that had no EE.

In addition, a higher proportion of alertness was observed in both the CE and PE conditions compared to the baseline conditions of my experiment.

These findings are consistent with the views of Carlstead and Shepherdson (1994) in their review of the effects of EE on the behaviour of zoo animals. They argue that an optimum level of arousal of captive animals by EE, shown by increased alertness and decreased aggression, is of more value than

an environment which is either barren or over-stimulating. The findings point towards subjects spending less time behaving agonistically with each other, and more time alert and interested in their surroundings when they are provided with either CE or PE, suggesting that both devices enriched the subjects' environment to a level that improved their welfare by increasing alertness whilst also decreasing aggressive behaviour.

4.5.2 Physical Condition (Expt. 1)

No significant differences were found in each subject's physical condition over the experimental conditions, leading to the assumption that while physical condition may be a good measure for longitudinal welfare, it is not very useful for picking up subtle differences between experimental conditions. However the findings showed that between individuals, subjects' physical condition correlated with stereotypic behaviour, with higher levels of stereotypic behaviour seen in pigeons in worse physical condition. While it is difficult to make directional causal inferences for this link, it is important that chronic physical defects such as foot problems, weight problems and parasites are promptly addressed as this may reduce incidences of stereotypic behaviour, and correspondingly if stereotypic behaviour is reduced, an improvement may be seen in physical health related to repetitive-pecking such as beak damage.

4.5.3 Stereotypic Behaviour (Expt. 1)

Stereotypic behaviour was found to be performed by eleven subjects and was observed in at least some individuals in all experimental conditions. The highest level of stereotypic behaviour was seen in the PE phase – one individual was observed engaging in stereotypic behaviour for over a third of the total observation time. The main stereotypic behaviour, which all stereotypic individuals carried out, was spot-pecking: repetitive pecking of a small hole in the wooden part of the dividing wall of the aviaries. Four individuals were observed repeatedly pecking the shelves and perches. One individual performed a stereotypic polydipsia behaviour, involving repeated pecking motions into water. One subject was observed pecking repeatedly in an empty water bath. One individual had developed a particularly severe stereotypic behaviour involving repeatedly air-pecking to the point of overbalancing off a perch, although this ceased after the introduction of CE into the enclosure and

was not observed again. All these stereotypic behaviours have repetitive pecking in common, and occurred particularly leading up to and after feeding time, suggesting that all were related to foraging frustration. Palya and Zacny (1980) conducted a series of experiments investigating the incidence and possible causes of spot-pecking in captive pigeons, and found a similar rate of occurrence to this study, and a similar time-frame of 4-7pm, which suggests that this behaviour is by no means unique to this group of subjects. Palya and Zacny (1980) linked spot-pecking to food deprivation, as increasing rations to free-feeding level eradicated the behaviour. As these subjects were not food-deprived in terms of the total amount they received, but instead received their ration by scatter-feeding at 4pm each day, there could be four reasons for the high incidence of stereotypic behaviour. Firstly, I found that physical condition was correlated with stereotypic behaviour, with subjects in worse physical condition performing more stereotypic behaviour. Birds that do not move as quickly and are displaced and supplanted by dominant individuals may not receive as much of a scatter-fed ration as fitter birds, which exacerbates the problem as a physically weaker bird is then less likely to be able to move quickly enough and defend its foraging area in order to get the increased ration it needs. Secondly, when Palya and Zacny (1980) increased the rations for subjects, there would have been an increase in total feeding time. Perhaps it is not the lower ration that causes foraging-frustrative stereotypic behaviour, rather it may be caused by the shorter successful foraging time. Thirdly, the specific feeding time of 4pm each day, in both this experiment and the one conducted by Palya & Zacny (1980), may cause husbandry-schedule stereotypic behaviour. A fourth possibility is illustrated by Latham and Mason (2010) concluding from their study of mice experiencing unenriched conditions after being raised with EE that animals experiencing a drop in enrichment can respond by performing higher levels of stereotypic behaviour than animals housed with no enrichment. The subjects in this experiment had previously been used in several cognitive experiments and had therefore experienced an enriched environment with the opportunity to develop cognitive skills that they would not have had in an unenriched environment. Their 'dormant' time before this experiment may have led to increased frustration, leading in turn to increased stereotypic behaviour. Latham and Mason (2010) advise against providing an enriching rearing environment if animals are likely to experience an

unenriched environment later in life. I suggest that all animals should be provided with continuous effective EE throughout their captive lives as an ethical consideration, and also because any data collected from animals known to engage in stereotypic behaviour may not be as valid and reliable as data from animals without this issue.

Stereotypic behaviour was significantly reduced in the CE condition compared to both the baseline conditions and the PE condition, suggesting that CE was effective in reducing the urge to perform stereotypic behaviour, which is in line with my hypothesis that stereotypic behaviour would be reduced in the CE condition. Fox *et al.* (2006) reported in their review of EE and its psychological effects that EE particularly benefits very anxious animals (i.e. those found to engage in high levels of stereotypic behaviour), and that more challenging EE (i.e. CE) was more effective than less challenging EE for these very anxious animals. In this study, there was no improvement found in stereotypic behaviour in the PE condition, which supports both the findings by Fox *et al.* (2006) and Mason *et al.* (2007), who argued that EE has had only partial success in eliminating stereotypic behaviour to date. The positive effect of CE on stereotypic behaviour suggests that it may be a more useful tool to develop than PE as an important aim of EE is to reduce stereotypic behaviour in captive animals. As Mason *et al.* (2007) propose in their review of stereotypic behaviour and EE, a stereotypic behaviour such as spot-pecking in birds would probably not be prevented by merely increasing cage size or complexity. As the underlying motivation for spot-pecking is likely to be feeding frustration, spot-pecking should be tackled with complex foraging opportunities.

Stereotypic behaviour may therefore be alleviated in this group of pigeons by making sure all individuals get an adequate share of food each day, by splitting up feeding over several bouts in a day, unpredictably if possible, by keeping a check on the physical condition of all birds and increasing the ration of any that require it, and by providing suitable CE for subjects to engage in behaviours that are frustrated by captivity.

4.5.4 Optimism (Expt. 1)

A significant difference in subjects' optimism scores was only found between the CE and baseline conditions. This is a very interesting finding for three reasons. Firstly, the large improvement in optimism during the CE

condition indicates that this type of enrichment is worthy of further study, especially as several other welfare measures also found it beneficial in this experiment. Secondly, there was no improvement in optimism in the PE condition, which leads to the conclusion that PE may not be as beneficial to welfare as thought by some, especially when evaluated in light of the finding that stereotypic behaviours did not decrease during the PE condition, and indeed seemed to increase for some subjects, although this needs rigorous testing with a range of PE devices and species. Thirdly, the fact that subjects' optimism scores fell repeatedly to baseline measures suggests that my method for measuring cognitive bias was effective over repeated presentations. The decision to reward the ambiguous probes in the cognitive bias task was a risky one, as subjects might have learned to interpret them all positively, but it was done in order to prevent the subjects from learning to interpret them all as negative over the conditions, a criticism raised by Doyle *et al.* (2010) and Brilot *et al.* (2010) regarding the more usual cognitive bias paradigm (e.g. Bateson and Matheson, 2007) where several repetitions of unreinforced ambiguous probes are presented for classification by subjects. The fact that optimism scores fluctuated between experimental conditions in this study suggests that subjects did not learn to interpret them all positively, and therefore it is suggested that a variable interval reward schedule of one reinforced probe to every three unreinforced probes is suitable for assessing subjects repeatedly when studying cognitive bias although this needs to be thoroughly tested.

Another very interesting finding was that optimism and stereotypic behaviour within individuals were not correlated with each other at all in any of the five experimental conditions, meaning that individuals that performed stereotypic behaviour, even at high levels, were not more likely to be more pessimistic than individuals that performed low levels of stereotypic behaviour, or even none at all. This finding is counterintuitive, as both stereotypic behaviour and optimism are measures of welfare, and they both improved overall during the application of CE and returned to baseline when the CE was removed. It counters the findings of both Brilot *et al.* (2010) who found that starlings that performed stereotypic somersaulting behaviour were more likely to interpret ambiguous stimuli pessimistically, and Pomerantz *et al.* (2012) who found that tufted capuchin monkeys that performed stereotypic head twirls were also more likely to interpret ambiguous stimuli negatively. However, Pomerantz

et al. (2012) also found that stereotypic pacing by the same tufted capuchins was unrelated to their judgement of ambiguous stimuli, and they suggest that interpreting all types of stereotypic behaviour as having the same value as welfare indicators may not be helpful, and that perhaps stereotypic behaviour caused by environmental restrictions such as pacing due to limited space (or in this experiment, stereotypic pecking due to reduced productive foraging time) can be interpreted as necessary energy-expenditure; and therefore, even though it still indicates lower welfare as the behaviours are unnatural and the environment should be altered to allow normal behaviour to take the place of stereotypic behaviour, the individuals performing the behaviours may not be experiencing as poor welfare as they appear to be. This would explain why in this experiment, subjects showed both reduced locomotion and stereotypic behaviour during the CE phase, as perhaps they could expend their energy budget more satisfactorily using the enrichment device.

Our findings suggest that even though some subjects did not perform stereotypic behaviour, their cognitive bias judgements indicated improvements in their welfare due to CE analogous to the improvements made to the welfare of the individuals that performed stereotypic behaviour; therefore the absence of stereotypic behaviour in subjects should not automatically be taken as an indication of better welfare without also taking into account their optimism. This supports the use of both stereotypic behaviour and optimism as tandem welfare indicators, as individuals may express their experience of captivity in different ways.

4.5.5 Use of Enrichment (Expt. 1)

The CED was used significantly more by subjects than the PED overall, leading to the conclusion that subjects were more motivated to use the CED despite the same quantity and quality of food rewards being available in both devices. This suggests that it may have been that subjects had to use their cognitive abilities to work out how to open the cups, and to choose the right colour cups to open, that made it more appealing for the subjects.

The CED and PED were used at similar rates for the first three days, but then subjects began to use the PED less, while becoming more engaged with the CED over time. After ten days, it appears it would be necessary to increase the complexity of the CED as subjects were beginning to use it for less time.

A path analysis was conducted on the use of the CED, which indicated that a large proportion of the increase in optimism was due to engagement with the device, rather than simply the device itself, meaning that the complexity of the design encouraged the pigeons to use it for longer, which in turn increased their optimism. There is presumably an optimum level of complexity that may differ between individuals. Experiment two examines the welfare effect of puzzles of different complexity utilising the same physical form as the CED.

The small size and robust design of the CED meant it was easy to use and survived the experiment well. For future studies, the cups could be turned or moved around, additional cups could be added, a 'signal' symbol could be placed in the middle to signify which colour cup to open, or different symbols could be applied to the tops of the cups, perhaps signifying how much food could be found in each cup. It is a very versatile piece of equipment, which is durable and easy to use, with plenty of scope for development. Although the PED was readily used by the subjects, it did not hold the subject's interest over the ten days, and was not found to affect welfare by any of the measures. It does not have any scope for furthering its enriching possibilities, as unlike the CED, as it cannot easily be altered or added to.

4.5.6 Predicting Welfare Improvement (Expt. 1)

A path analysis indicated that the type of enrichment (CE or PE) had a large effect on how long subjects used each device, which in turn had a smaller effect on their optimism scores. This suggests that an important aim for any enrichment is to get subjects interacting with it as much as possible in order to raise welfare.

Use of enrichment did not have any predictive value for individual stereotypic behaviour, which is similar to the findings of Dallaire *et al.* (2012) with American mink, although for the group as a whole, introduction of the CED reduced levels of stereotypic behaviour. Perhaps being able to observe others interacting with a CE device may be enough to raise welfare in some individuals as Platt and Nowak (1997) found with rhesus monkeys, and Tarou *et al.* (2004) found with orang-utans. 'Use of enrichment' should therefore include 'watching others interact with enrichment'.

4.5.7 Order Effects (Expt. 1)

The possible effects of the fixed order schedule of the experiment seem to have been minimal, as shown by the variability in the measures taken. For example, stereotypic behaviour reduced in the CE condition then rose again in the second and third baseline condition, while optimism as measured by the cognitive bias task rose during the CE condition, but then returned during the second and third baseline condition to levels previously seen in the first baseline condition, and so on. Therefore, a repeated measures design with a baseline measurement condition before and after each experimental manipulation is not thought to be a severe limitation of the experiment, and indeed it can be beneficial to see how the same subjects respond to different manipulations, with each individual serving as their own control, and at the same time avoid comparing the results of different groups of subjects that may have inherent differences.

4.5.8 Stereotypic behaviour (Expt. 2)

The amount of stereotypic behaviour subjects performed reduced significantly between baseline one and the CE condition. However, the type of enrichment, either high or low complexity, had no effect on stereotypic behaviour. This means that although the subjects benefitted significantly from access to CE, the level of complexity could not account for the differences.

A significant reduction of stereotypic behaviour is an important improvement, which suggests that CE should be made a priority for animal welfare research. Mason *et al.* (2007) argue that the aim of any intervention should be to eradicate stereotypic behaviour. In the CE condition of the experiment, stereotypic behaviour was indeed eradicated in seven of the subjects that previously performed stereotypic behaviour. In the second baseline measure, two of those subjects had relapsed into stereotypic behaviour, and the mean had begun to rise again, in keeping with Latham and Mason (2010) who found that mice experiencing barren conditions after an enriched environment showed higher stereotypic behaviour than mice that were just kept in barren conditions. This suggests that CE should be sustained over the whole life of the captive animal in order to (a) prevent stereotypic behaviour from taking hold, (b) to keep any stereotypic behaviour that does develop to a

minimum and (c) prevent any 'rebound' stereotypic behaviour from occurring due to frustration from having previously experienced CE.

4.5.9 Optimism (Expt. 2)

Optimism scores for subjects increased significantly between measures taken in baseline one and the CE condition, before dropping again in baseline two. There was a significant interaction between optimism and enrichment group. This means that optimism varied significantly according to whether subjects experienced the low or high complexity enrichment, with a larger effect in the low-complexity group, and both enrichment types raised subjects' optimism scores to approximately the same level. Despite subjects being assigned to the high and low complexity groups randomly using researchrandomiser.org, the low-complexity group were notably (but not significantly) less optimistic in the baseline phase than the high-complexity group. The data were not analysed until the experiment was completed, thus missing this anomaly. A better procedure might have been to match subjects in the two groups for initial optimism score.

The high-complexity CE device was based on a simple discrimination between two colours where one was rewarded and the other was not. The low-complexity CE device was of the same structural design but all the food-containers were the same colour. There is a possibility that it was the 'surprise' element of not being able to guess the positive from the negative stimuli that raised the subjects' optimism when it came to the cognitive bias task. There is also a possibility that subjects in the low complexity CE device group were using acoustic signals to figure out which containers had food in them – when pecked the containers made a slight rattle if they were full. This was only noticed when listening back to the recordings, and was therefore not remedied. The same effect was also possible in the high-complexity CE device group but colour possibly held more value as a cue in the noisy room. This could be investigated further using experiments involving other sensory stimuli.

Consistent with findings in experiment one, stereotypic behaviour and optimism were not correlated with each other at all, meaning that individuals showing higher levels of stereotypic behaviour were not more likely to have a pessimistic judgement bias. This is unlike the findings of Brilot *et al.* (2010) and Pomerantz *et al.* (2012), both of which highlighted a link between stereotypic

behaviour and optimism. However, Pomerantz *et al.* (2012) argued that not all stereotypic behaviours are based on the same underlying motivations, and that some may not represent an experience of lower welfare for the individual, instead providing an outlet for energy expenditure. This demonstrates that although both stereotypic behaviour and optimism are welfare indicators, perhaps stereotypic behaviour alone cannot be relied upon to indicate the animals' emotional experience of captivity, and both measures should be used in tandem in order to gain a fuller picture of welfare.

4.5.10 Use of Enrichment (Expt. 2)

Subjects did not begin to interact significantly with either enrichment until day five, whereas in experiment one, during which EE was presented in their group enclosure, subjects began to interact with enrichment devices on day one. The reason for this could be that out of their home environment, isolated from other individuals, subjects felt less safe and therefore were less interested in interacting with EE. Another possibility is that, as Giraldeau and Lefebvre (1987) found with a group of pigeons, there were one or two producers and the rest of the group were scroungers, and although scroungers can learn foraging tasks, they prefer to leave it to the producers if possible and also learn novel foraging tasks more slowly when there is no demonstrator, as there was in experiment one as the enrichment devices were presented to the whole group.

However, all subjects interacted with the enrichment devices eventually. Subjects were categorized as low or high enrichment users by whether they pecked at the enrichment device more or less than the median number of pecks per session. Subjects in the high-use category showed a larger increase in optimism than the low-use category. As the high-users' optimism scores started off similar to those in the low-use category, and returned to the same level in the second baseline measure, it is more likely that the use of enrichment affected their optimism, rather than it being that more optimistic individuals were more likely to use the enrichment more; if this had been the case I would have expected higher starting scores in baseline one. This implies that, in keeping with the suggestions from experiment one, high engagement with enrichment should be encouraged through design and rewards in order to maximise welfare improvements.

Over time, use of both CE devices increased significantly. The high-complexity group used their CE device significantly more however, with a steady increase over time, and no evidence of habituation. The low-complexity CE device conversely seemed to begin to lose value towards the end of the experiment with gradually less use by subjects. The benefit to increasing the complexity of a CE device may be that higher complexity increases the amount that subjects interact with it, with less habituation over time than with less complex enrichment. Further investigation is needed to assess the value of each enrichment type over a longer period of time, but it is reasonable to assume that a more complex device would be habituated to less and would encourage more use over time.

The enrichment devices themselves are both capable of becoming more complex – the LCCED could have more cups added with the same number of rewarded ones, thus decreasing the chance of success and increasing the time spent using the device to obtain the same ration, while the HCCED could have more cups but also a second discrimination or even more, perhaps an if/then contingency could be added (e.g. IF there is a blue square in the middle THEN the light green cups are rewarded). The value of this device design is the relative ease with which it can be utilized by technicians and keepers, and as both group and individual enrichment showed benefits in terms of reducing stereotypic behaviour and increasing optimism in captive animals, whichever style of housing subjects have can be easily, cheaply and quickly enriched. This is therefore an important development in captive animal enrichment and welfare research.

4.5.11 Group versus Individual Enrichment Presentation (Expt. 2)

I expected that either group enrichment would show greater welfare improvements, possibly because of a feeling of safety in numbers and/or opportunities for social interaction and facilitation; or that individual enrichment show greater welfare increases by creating a safe environment without dominant individuals controlling access to the devices, and allowing for individual learning.

The incidence of stereotypic behaviour was significantly higher in the first baseline measure for experiment two compared with experiment one. However, both group and individually presented CE reduced stereotypic behaviour to a

similar level, suggesting that perhaps there is a low set-point that can be reached when subjects experience CE. Stereotypic behaviour was measured in the same way for both experiment one and two, with all occurrences being recorded between the hours of 2-7pm on a single day during each condition. The reason for higher prevalence of stereotypic behaviour in subjects at the beginning of experiment two compared with experiment one could be to do with having previously experienced EE, particularly CE, in experiment one, then having a dormant period of seven months where subjects had no EE above their normal amount, leading to a resurgence in stereotypic behaviour above what was previously observed in the baseline phase of experiment one, as subjects had finished a cognitive experiment where they experienced daily training on a cognitive task only two months previous to starting experiment one.

The benefit of carrying out more than one welfare measure is highlighted by the complications seen in the measurement of stereotypic behaviour across experiment one and two. Both group- and individually-presented CE increased optimism in subjects by approximately 25%, suggesting that it is the enrichment itself rather than the setting that increases optimism in subjects, as the cognitive bias test was identical in both experiments. Optimism measures for baseline conditions were very similar to each other, as were CE condition measures, suggesting that optimism is a very stable and reliable measure of animal welfare. A bonus of this similarity in the amount that optimism increased in both experiments is the support it gives to my design alteration of the cognitive bias task – a variable interval reward schedule was used to prevent subjects learning a negative association to the ambiguous probe stimuli. The findings indicate that subjects have continued to view the ambiguous stimuli as intended – ambiguous. If subjects had learned a negative association for the probe stimuli over time a more pessimistic score would have been expected for both phases of experiment two compared with experiment one.

Manteuffel *et al.* (2009) discussed the difficulties in presenting group-housed animals with CE, and suggested that it is not beneficial to isolate group-housed animals who may become stressed by the separation. Instead, they suggested ways in which the enrichment can be presented to individuals without isolating them, such as providing several spatially separated automated enrichment devices that can 'call' individuals using an identifying tone. While

this may be beneficial for reducing conflict and aggression in animals that may fight over access to enrichment devices, in experiment one no such behaviour was seen. Perhaps an alternative to individually-presented enrichment could be a large enough CE device that all individuals can participate simultaneously, allowing social learning, facilitation and interaction to occur which may further enrich the enrichment experience. Csatadi *et al.* (2008) looked at the effects of introducing enrichment to a group of captive bonobos which had developed a negative behaviour (stealing and rough handling of an infant). They noted that enrichments that worked at a group level were particularly effective, while enrichments which only one individual could dominate were less effective. This suggests that in groups with marked dominance structures, any enrichment should be large and complex, or numerous enough for all the animals in a group-housed unit to use at once without increasing conflict.

It appears therefore that both group- and individually-presented CE is beneficial, at least to captive pigeons, even though subjects did not begin to interact with the devices until day five in the individually-presented phase, and mode of presentation could therefore be based on the animal's home environment and other needs – for example if animals are group-housed, group-housed enrichment could be provided, while if animals are individually-housed, appropriately-sized individual CE devices could be used. Alternatively, if individually-housed animals are given time to socialize with conspecifics, CE could be incorporated as there seems to be no effect of dominance on use of or benefit from CE at least in the case of pigeons (see section on dominance below).

One criticism of the experiment is that the measurement of enrichment use differed between the two experiments, making direct comparisons of enrichment use difficult. The reason I chose to switch from a one-zero sampling method in experiment one, to recording individual pecks using an event-logger in experiment two, was that it enabled increased accuracy – a one-zero approach only recorded whether subjects were seen interacting with the enrichment at all during each minute. In hindsight keeping the methodology identical would have facilitated easier comparisons.

4.5.12 Dominance (Expt. 2)

Dominance was investigated by observing agonistic interactions and constructing a hierarchy by calculating David's Scores for each subject, then ranking subjects during each experimental condition. The hierarchy was found to be strong, stable and linear over time.

However, subjects' DS increased significantly during the CE condition of the experiment. This means that there were more agonistic encounters overall in that time, probably caused by the social upheaval brought about by daily time out of the home aviary in order to engage with the CE in individual cages. This was not considered in the design of the experiment, and should be taken into account when designing enrichment for normally group-housed species, and also considered when using animals for cognitive experimentation, where separation of individuals is common practice.

I expected that either lower-ranking pigeons would be slower at learning and engage less with the CE device than higher ranking pigeons even when presented with enrichment individually, consistent with findings in goats by Baymann *et al.* (2007); or that dominance, learning ability and engagement with the enrichment would not be correlated, as found by Giraldeau and Lefebvre (1987) and Croney *et al.* (2007). Surprisingly, in this experiment, subjects with a higher DS, meaning the more dominant individuals, also took longer to learn the discrimination task. Although the learning task took place seven months prior to dominance measurement, the hierarchy remained fairly stable over the three measurements taken in experiment two. Perhaps the more dominant individuals could not concentrate as well on the learning task because of added pressures on them to defend their position in the hierarchy during the time of upheaval in experiment one. This could be a potential confound in cognitive experiments if the learning ability of dominant individuals is compromised by instability within their group due to individual training sessions necessitating removal from the group.

Dominance did not have any relationship with the amount subjects interacted with the enrichment in either group or individually presented enrichment. This is consistent with Elmore *et al.* (2011) who found that both dominant and subordinate sows showed welfare increases in an enriched environment, even though subordinate sows had to use enrichments at off-peak times and were displaced from it more often.

Dominance was also uncorrelated with stereotypic behaviour and optimism in this study, suggesting that the strong linear hierarchy did not induce any measured welfare benefits or detriments. However, as dominance is also often correlated with physical condition, and stereotypic behaviour and physical condition had been found previously to correlate in this group of pigeons (experiment one, this thesis), perhaps one observation session of three hours in each condition was not enough to gain a full picture of dominance and its links with welfare.

4.5.13 Predicting Welfare Improvement (Expt. 2)

A path analysis indicated that, as in experiment one, high use of enrichment predicted higher optimism. Whether the enrichment device was low- or high-complexity, subjects' learning ability and dominance had no predictive value for optimism, while stereotypic behaviour could not be predicted accurately from any measure. This is similar to the findings of Dallaire *et al.* (2012) who also found that stereotypic behaviour could not be predicted through subjects' use of enrichment.

4.5.14 Conclusions (Expt. 1)

The provision of EE that reaches the goal of improving captive animal welfare is hampered by problems with both the definition of types of enrichment and measurement of welfare. CE is a new area of research, which is gaining respect in the field of animal welfare, and which has been found by this study to provide welfare benefits to group-housed pigeons, by reducing stereotypic behaviour and autogrooming, and increasing optimism. This is consistent with the theory that optimism represents a positive affect as it corresponded to the reduction of stereotypic behaviour and autogrooming, and taken together these measures suggest subjects experienced a welfare improvement during the CE condition.

Both CE and PE were found to reduce agonistic behaviour and increase alertness. However, PE was found by this study not to provide any measureable welfare benefits to group-housed pigeons in terms of reducing stereotypic behaviour or increasing optimism. Subjects used the CED significantly more than the PED during the time the devices were available. Overall, the findings suggest that CE may be more useful for improving captive animal welfare than

PE, and the effects of CE on captive animal welfare is worthy of further investigation.

4.5.15 Conclusions (Expt. 2)

Both low and high complexity CE were associated with significant welfare improvements in terms of reducing stereotypic behaviour and increasing optimism in subjects in this experiment. In answer to the question posed at the outset of this experiment, whether welfare improvements associated with CE are due to the fact that food was more difficult to extract from it than the PE device; or that the puzzle stimulated their interest for longer, both devices reduced stereotypic behaviour to a similar level, but higher complexity of the device was associated with increased use over time, which in turn increased optimism. So, providing an engaging cognitive challenge that encourages high use through the provision of rewards, that resists habituation over time through the capacity for alteration, seems to be of most benefit to captive animals.

Both group and individual presentation of CE seem beneficial to captive pigeons, but this is likely to differ between species with different social structures and agonistic tendencies. If animals are group housed, CE should be large or numerous enough to facilitate participation by all members of a group. If animals are housed individually, CE could either be presented in their home environment or if facilities allow, group-presented CE may allow for a richer experience. Dominance and learning ability were correlated in this experiment with lower learning abilities seen in higher-ranking individuals. This could have been a side-effect of the daily time out of the home aviary and subsequent rise in agonistic encounters. This should be considered when making the decision to present enrichment to the group or individually. On balance, as a consideration to technicians and other care-givers, group-housed enrichment is easier to provide to group-housed animals if all other factors permit.

This experiment only provided subjects with enrichment for ten days, so further investigation is needed to assess the value of this type of enrichment over a longer period of time.

5.0 Chapter Three

The design and implementation of a cognitive enrichment device known as the DogBox

Louise N. Millar*, Stephen E.G. Lea, Lisa A. Leaver

**Corresponding author: Louise N. Millar, University of Exeter, School of Psychology, Washington Singer Laboratories, Streatham Campus, Perry Road, EX4 4QG. Email: L.N.Millar@exeter.ac.uk.*

Abstract

Cognitive enrichment (CE) has been found to improve the welfare of captive pigeons (experiment one and two, this thesis) as well as other captive animals in zoos (Clark, 2011) and farms (e.g. Manteuffel *et al.*, 2009). The effects of CE upon the welfare of these animals have been assessed by measuring activity budget changes, stereotypic behaviour, 'optimism' via cognitive bias tasks (e.g. Bateson and Matheson, 2007), glucocorticoids and other physiological measures and behavioural reactivity tests such as novel object approach. However, CE is time-consuming to provide to captive animals as it is important to keep the cognitive challenge at a level for each individual that prevents boredom, anxiety or habituation through over-familiarity. What is needed is a way to provide individual animals with CE without the need for a human trainer to be present, and for the task to be easily programmable to tailor the tasks to the individual. Any implementation needs to be easy to present to animals in their home environment and quick to move from one place to another, without requiring too much time-consuming preparation or expert knowledge. We therefore designed a device which had the features of being small, portable, easy to use, flexible and easily programmable, easy for animals to use, and most importantly effective in improving the welfare of captive animals.

*Keywords: Operant training, animal welfare, cognitive enrichment, stereotypic behaviour, learning ability, dogs, *Canis familiaris*, captivity, isolation.*

5.1 Introduction

Currently, environmental enrichment (EE) is often thought of as expensive, time-consuming and a hassle to implement by human caregivers, whilst also being quickly habituated to, and without increasing welfare significantly, while welfare measurement problems also make comparisons between research studies arduous (Newberry, 1995; Young, 2003; Fox *et al.*, 2006; Meehan and Mench, 2007; Azevedo *et al.*, 2007; Adams, 2007; Manteuffel *et al.* 2009; Clark, 2011; Franks, 2012).

There is a movement towards cognitive enrichment (CE) as an important welfare consideration for captive animals (Meehan and Mench, 2007; Manteuffel *et al.* 2009; Clark, 2011; Franks, 2012). CE can be defined as enrichment of the lives of captive animals by challenging their cognitive abilities through applications such as puzzles and positive reinforcement training.

In their review of CE techniques for farm animals, Manteuffel *et al.* (2009) describe how successfully coping with a cognitive challenge via integrated operant devices in their home environment improves the welfare of captive animals by reducing boredom and frustration.

Using computer assisted CE is an important step forward as animals have been found to respond well to computer-generated stimuli presented on screens in studies of motivation and cognitive ability (Rumbaugh *et al.*, 1989; Washburn *et al.*, 1989; 1990, 1991; Wilkie *et al.*, 1994; Leighty and Fragaszy, 2003; Wills *et al.*, 2009).

Fagot and Paleressompoulle (2009) provided baboons with access to computerised cognitive testing equipment twenty-four hours a day in their home environment. The baboons demonstrated that they could solve complicated cognitive puzzles, and indicated their desire to partake by their high level of voluntary participation.

Platt and Novak (1997) found that rhesus macaques were able to control a cursor via a joystick to obtain food in their home environment, and showed higher activity levels indicative of improved welfare. They did not seem to habituate to the device over time, possibly because the task varied in difficulty suited to the abilities of the user, helping them to achieve 'flow' and reducing the risk of apathy or anxiety. They also found no evidence of increased aggression.

Mandell and Sackett (2008) reported on the use of computer touchscreen procedures with infant pigtail monkeys and found that even very young monkeys could use a touchscreen meaningfully and learn operant tasks successfully, suggesting that while there can be problems getting subjects to use a joystick or keys to interact within a computer environment, a touchscreen removes physical barriers from the procedure and enables subjects to engage in cognitive tasks more easily.

Cognitive research into the abilities of animals has demonstrated that pigeons (Wilkie *et al.*, 1994; Wills *et al.*, 2009) and dogs (Range *et al.*, 2008) can and do interact with a computer touchscreen by learning discrimination tasks in order to gain rewards.

Common problems with current designs of operant learning devices are (a) the device is in a fixed place such as a laboratory testing room and animals have to be taken to a separate area in order to interact with it (e.g. Wills *et al.*, 2009); (b) the device is built into the environment in which the animal lives, making quick changes to the programming and swaps to other animals' cages difficult, and/or (c) the equipment is so complicated and fiddly that only trained technicians can operate it successfully, and maintenance is expensive and time-consuming.

Due to the problems associated with captive animal welfare, there is a need for a portable, simple, easy to use, flexible, habituation-resistant, truly enriching enrichment device.

5.2 DogBox Development

The DogBox was created in 2008 by L. Millar after working in the Animal Behaviour department of the University of Exeter, UK, with laboratory pigeons that were used in cognitive experiments in specially adapted Skinner boxes, complete with infra-red touchscreens (pigeons cannot use capacitive screens due to their beaks being non-conductive, and pressure-sensitive ones are often too resistant for their light touch) and automated reward dispensers. The pigeons had to be manually caught, weighed and placed in the boxes, then removed and weighed again at the end of the session. Simultaneously, cognitive experiments were being carried out with squirrels in an adjacent room, and the squirrels were so highly motivated to access the cognitive testing room that they broke through a strong metal door to get there. It seemed that the

animals all benefitted (evidenced by increased physical health and less stereotypic behaviour than when not being used for experiments, personal observation) from the CE provided by the experiments, but having to be physically transported to the devices and preventing access when the animals were motivated to participate was not optimising welfare.

5.3 DogBox Design

The DogBox device is essentially a small (approximately 40cm high by 40cm long by 30cm wide, and approximately 6kg), portable (handles and a strap for carrying by one person), inverted Skinner box, not a novel concept in itself (See Range *et al.* 2008), but several aspects of the DogBox combine to make it a very useful device.

Fig. 1 shows the front of the DogBox, an iPad™ flush to the frame with the 'home' button inaccessible (important as it prevents animals accessing the main menu of the iPad™), and a small chute through which rewards are dispensed. This is what subjects see when the device is attached to the front of their cage using the hooks at the top – usually mesh in the case of dogs in kennels. The second picture is of the inside of the device, showing the round revolving reward dispenser, which is designed to be filled with up to 40 Barker and Barker Little Liver Treats®. These treats were chosen as they are small, round tablets that do not clog up and jam the device, and are only 0.2g each. The total amount of up to 60 rewards available from the DogBox in a training session (if the device is refilled) fits the guidelines given by www.barkerandbarkertreats.co.uk that between 60-150 Little Liver Treats® may be given per day without altering dogs' calorie intake - 60 treats at 0.2g each totals at 12g, around the same weight as one Pedigree Chum Schmacko®, another leading treat brand.

The recommended daily dose of these treats without having to adjust calorie intake is between 60-200 treats, making them very suited to reinforcement of the DogBox. The treat dispenser is triggered by a sound-to-electric-pulse device, in turn triggered by the iPad's headphone socket when a S+ is touched and makes a sound. A headphone splitter and portable speaker allows for the trigger to happen along with the subject being able to hear the sound of the S+ for clarification, similar to the effect of a 'clicker' when an animal is trained by a human trainer. There is also a remote trigger in case of

needing to 'shape' a subject – meaning that if an animal does not understand the link between touching the screen and getting a reward for example, a remote trigger of the reward dispenser each time an animal gets closer to the DogBox, then the screen, then touches the screen, then touches the S+ for example, will 'shape' the desired behaviour. There is also a chargeable battery, a sound-to-electronic-pulse converter device, a four-way plug adaptor and a panel to access the iPad™.

The DogBox is purposely fairly low-tech compared to the equipment often found in cognitive testing laboratories, having only two parts that were specially made for the device – the reward dispenser and the solenoid that moves the reward dispenser along one space upon receiving an electric pulse. This means that the device can be easily used, fixed and adapted depending on the situation and the skill level of the staff at the animal facility, unlike most high-tech enrichment and cognitive experimentation devices currently being implemented at zoos and laboratories worldwide.



Fig. 1: The DogBox. The front view (left picture) shows the iPad™ in situ displaying a moving stimulus, the reward chute at bottom centre, and the hooks at the top used to attach it to the front of an animal enclosure such as a dog kennel. The inside view (right picture) shows the electronic reward dispenser with space for 40 small liver flavoured dog treats (Barker and Barker little liver treats™). Total of 60 rewards used in experiment 4, with dispenser refilled part-way through each session.

5.4 Programming

Programming for the iPad™ was done using the iOS SDK 2010, version 4.0. The language used to create an ‘app’ (as programs built to run on apple handheld devices are known) is Objective-C, a relative of C and C++. The environment used to write the programmes is known as XCode. Developers are permitted to develop apps and test them with one device, but must be registered and approved before releasing apps on the ‘App Store’ to the general public.

The programme that was designed for and tested in experiment four had the following properties: There was a black background screen, upon which one large S+ (either an orange star or a blue circle, 10cm in diameter) was then presented in the centre to begin the session. The colour scheme was chosen on the basis of a previous experiment (Wills *et al.*, 2009) as blue and orange had

similar salience for squirrels, known to have a dichromatic visual system similar to domestic dogs. The shapes of the stimuli were also based on those used by Wills *et al.* (2009), as they had the same overall number of pixels and were the same brightness. Counterbalancing was used routinely during implementation of the programme in experiment four, and no evidence of one stimulus having higher salience than the other was seen. The S+ stayed on the screen until touched by a subject once, upon which the programme triggered a sound, which was converted to an electrical pulse, which in turn triggered the reward dispenser, which then dispensed a reward down the delivery chute towards the subject. Only the shapes registered touch, therefore touching the black background did not result in any reward or punishment.

Stage One:

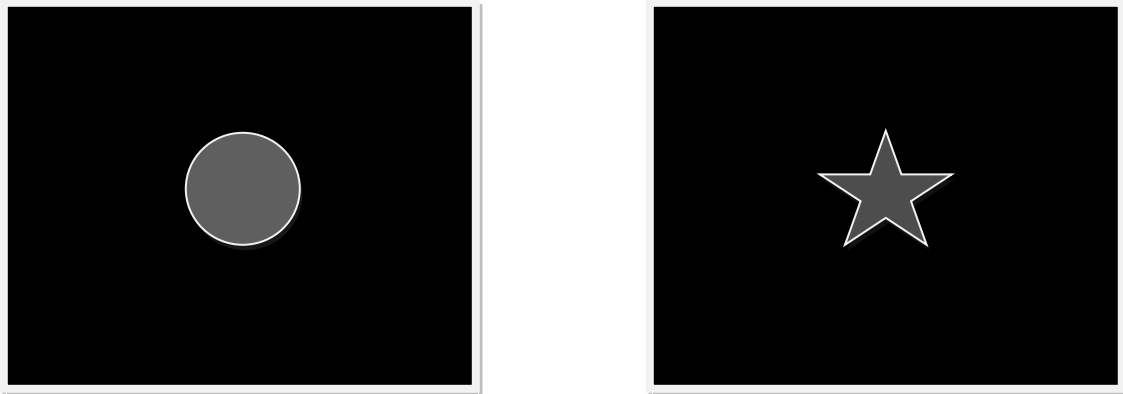


Fig. 2: S+ blue circle or orange star, central, static, 10cm diameter, when touched a beep is sounded which triggers the reward dispenser. When screen is devoid of S+ it is refreshed with another S+ (or batch of S+ and S- in later stages) until end of session which is 20 minutes long OR 60 rewarded S+ whichever is sooner.

Stage Two:

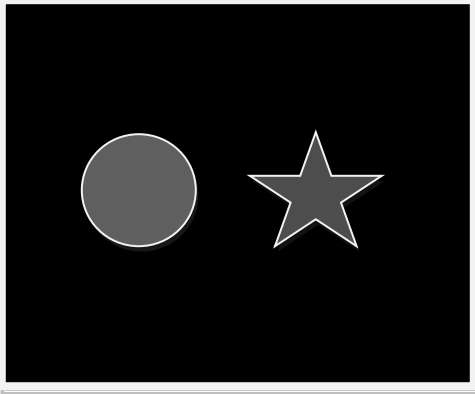


Fig. 3: S+ paired with S- in a randomised place on the screen, i.e. not just left/right as in example above. If S+ is touched, reward as before. If S- is touched, S- disappears but no other action occurs until only S+ remain. If black background is touched no action occurs.

Stage Three:

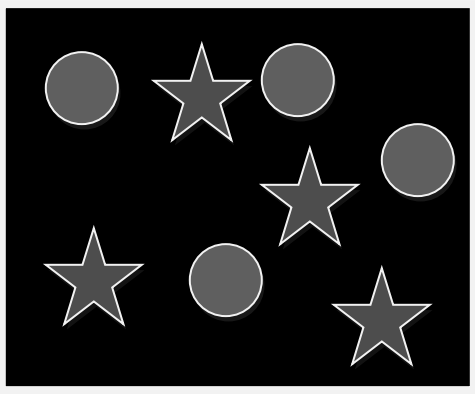


Fig. 4: Several S+ and S- presented together, reward schedule as before. Each stimulus is now 6cm in diameter.

Stage Four and Beyond:

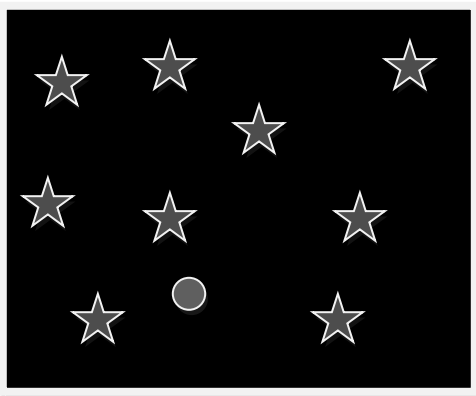


Fig. 5: S+ shown here as blue circle, total number of S+ available on screen at one time gradually reduces to one (by increments of 1 per session at criterion or above), while S- become more numerous by the same increment. The stimuli also became smaller (down to 1cm in diameter in 1cm increments per session at criterion or above) so as to introduce both increasingly difficult visual search and skilful touch needed. The stimuli then begin to move, in increments of 0.5cm/s per session at criterion or above, increasing in speed to 5cm/s until 'catching' the S+ is needed, similar to a cat chasing a laser pen dot.

This is a basic CE programme, designed to hold the interest of animals for a short time such as a week of daily training sessions. It can be added to and changed however, which reduces the risk of habituation setting in over time. In experiment four, eleven dogs were given access to the DogBox showing this programme, and in five successive presentations of 20 minutes per day all dogs reached stage three with no evidence of habituation.

Other program ideas, based on common cognitive research paradigms, include match-to-sample, categorisations, if-then contingencies, and a cognitive bias testing scheme (see Bateson and Matheson, 2007, also experiment one, two and four this thesis) could be implemented, thus potentially improving and testing welfare at the same time. Different reward contingencies could also be introduced to increase challenge and reduce food reward provision.

5.5 Key Benefits

Effective at providing CE to captive animals

The DogBox has been found in experiment four to be effective in significantly improving the welfare of kennelled dogs by reducing stereotypic behaviour and increasing optimism (see experiment one, two, four this thesis). The device needs to be tested with a wide variety of animals and settings to validate it properly but initial findings are very promising.

Adaptable and Flexible

The cognitive puzzles that can be programmed are very flexible, so animals can move up in levels of complexity or be presented with different concepts to suit the setting, the species and the individual from laboratory mice to zoo animals such as primates, ungulates and parrots, to farm animals such as sheep, pigs and goats, to companion animals such as dogs in kennel settings as well as in the domestic home environment. Important to note: the iPad™ has a capacitive screen therefore cannot be used with some animals that will use a non-conductive part of their body to touch the screen such as the beaks of birds – a device with a resistive screen should be used with these animals; there are resistive tablets available that will fit the device.

With a little adaptation to the reward dispenser, rewards could vary from monkey-chow to nuts to dog treats to sugar balls to hay pellets depending on the animal. The device could even be used with non-verbal or infant humans, who could be rewarded with sweets, tokens or other appealing items.

A randomised timer could be built in to the programmes so that a new task could be signalled by an attention stimulus at a random time during daylight hours in order to prevent schedule stereotypic behaviour from developing, as this is a risk with any rigidly scheduled system (e.g. Bloomsmith and Lambeth, 1995; Vickery and Mason, 2004).

Multi-purpose

It is proposed although not yet tested that the device could simultaneously be used to deliver CE, test cognitive abilities and measure welfare by presenting a cognitive bias task (e.g. experiment one and two, this

thesis) as an enriching puzzle. Experiment four in this thesis documents an example of a cognitive bias testing procedure that would work well as a digital version.

Communication and control are other possibilities, by using 'yes/no' or 'this/that' discriminations. Animals could be taught to press an 'attention' button in order to gain the attention of keepers, and could gain control over their environment: for example by indicating choices on the screen, two food options could be chosen between; the lights could be dimmed or brightened; a door could be shut or opened.

The device could also be used as a video screen requiring no operant input from the animal, for example to provide visual and/or aural enrichment as an alternative to CE when the animal's allotted rewards have been used for the day.

Accessible

The screen is a commercially available iPad™, meaning that many people will already own one, and can simply slot it into the DogBox, and special programmes can be written and provided for use in the DogBox. The programmes do not require any training of the animal by the owner/keeper in order to work, and the animal learns at their own pace.

Useful Research Tool

An on-screen, printable or exportable results summary could be prepared which explains what has been found in terms of enrichment use, learning and welfare measures clearly for interpretation by both the researcher and layperson. This is yet to be implemented but obtaining results datasets from operant testing devices in laboratories is common practice and this property should be possible.

Safe

Animals only come into contact with the front part of the DogBox, the screen of which is protected by a strong transparent cover. There is no risk of electric shock when used properly, i.e. accessed by the animal through the mesh of their home enclosure, to which it easily attaches therefore requiring no

animal/human contact or risk of escape. It can also be used free-standing for companion animals if appropriate supervision is given.

Portable

The device is small, light, needs no external power (runs off a chargeable battery), has handles and a strap and can be carried easily by one person.

Quiet and unobtrusive

The reward device is triggered using an electric pulse, which is silent and will not disturb other animals or scare subjects. There is an option for a sound to be heard when touching S+ to facilitate learning if needed.

5.6 Conclusions

The DogBox has a place in both welfare research and welfare provision, for all cognitively and physically able animals in zoos, laboratories, farms and domestic homes. It is easy to use, portable, habituation-resistant and has been shown to be effective at providing a CE task to kennelled dogs that measurably improved their welfare in terms of reduced stereotypic behaviour and increased optimism (experiment four, this thesis).

More research is now needed to test the value of the DogBox as a CE device for a range of captive animals in various settings, to develop more CE programmes, and to investigate the possibility of presenting cognitive bias tasks via the DogBox as both CE and welfare measurement.

6.0 Introduction to Experiments Three and Four

In chapter four I discuss experiments three and four carried out with a group of fifty-eight individually kennelled dogs at two rescue shelters. Cognitive enrichment (CE) has been found to be beneficial to the welfare of farm (Manteuffel *et al.*, 2009) and zoo (Clark, 2011) animals, as well as laboratory pigeons (experiment one and two, this thesis).

The CE device known as the DogBox that I developed for use with the dogs in order to provide effective, adaptable, flexible, portable CE was described in chapter three. I found the device to be capable of providing effective CE to the dogs involved in experiment four, and with further development would be suitable to use as a CE device for most captive animals in most settings.

Through experiments one and two, and in designing the DogBox device, I began to develop a useable CE method and welfare measurement system, which I then sought to test with other captive animals. I chose dogs as a second study species as they are sociable, intelligent animals that spend a lot of time in isolation in kennels, which is known to cause welfare problems for the dogs. I predicted that providing CE to individually housed dogs might alleviate some of the common welfare problems such as stereotypic behaviour and negative affect for subjects. I measured activity budget, stereotypic behaviour and optimism (based on my previous cognitive bias paradigm, see chapter two) and found that dog welfare is not optimised by being housed in individual kennels (experiment three), and that CE did improve some of the welfare parameters measured (experiment four). This suggests that CE, which has already been found to be useful for captive farm, zoo and laboratory animals, may also improve the welfare of kennelled domestic dogs.

Finally, I further enhanced the cognitive bias paradigm based on Bateson and Matheson (2007), which I first modified for experiments one and two, by altering the reward schedule. In experiment four I additionally altered the presentation of stimuli and added an 'advance key' in order to clarify answers further, which seemed to work well for the purpose, and is fully described in chapter four.

6.1 Chapter Four: Experiment Three and Four

In kennelled dogs (*Canis familiaris*), stereotypic behaviour and learning ability vary according to length of stay, and cognitive enrichment improves welfare

Louise N. Millar*, Stephen E.G. Lea, Lisa A. Leaver

**Corresponding author: Louise N. Millar, University of Exeter, School of Psychology, Washington Singer Laboratories, Streatham Campus, Perry Road, EX4 4QG. Email: L.N.Millar@exeter.ac.uk.*

Abstract

Social animals such as dogs that are kept in isolation are generally considered to have lower welfare compared to animals with access to conspecifics, and may develop undesirable behaviours such as stereotypic behaviour through lack of socialization opportunities with humans and conspecifics, under-stimulation and restriction of natural behaviour. Stereotypic behaviours have been found to correlate with lower learning abilities in other animals, and may be detrimental to re-homing dogs because of the undesirability of stereotypic behaviour in a family pet (sometimes stereotypic behaviours persist despite an improvement in situation – a welfare scar), and because a dog that performs stereotypic behaviour may be more difficult to train.

The behaviour and trainability of fifty-eight dogs in two animal shelters in the UK were studied over several weeks. The key findings were that dogs housed in one shelter played with toys in their kennels more, and were taken out of their kennels for exercise and social interaction less, than dogs housed at another shelter. Dogs that had been in a shelter for more than a week were less sociable and engaged in more stereotypic behaviour than dogs that had been in for a week or less. Dogs that had been at a shelter for longer than a week, and were showing signs of stereotypic behaviour, took longer to learn a cognitive task than short-stay dogs that did not perform stereotypic behaviour, suggesting a link between length of stay, stereotypic behaviour and reduced learning

ability. It is possible that a kennel stay adversely affects dogs, although the reasons for the dogs being relinquished to the shelters also needs to be taken into account.

The provision of suitable, effective environmental enrichment (EE) for captive animals is an important issue in animal welfare, as it has been shown to improve the neurobiology, physiology and behaviour of captive animals. Cognitive enrichment (CE), or enrichment which requires animals to use their cognitive abilities, has been found to have particularly significant effects on captive animal welfare (experiments one and two in this thesis). Social animals such as dogs that are kept in isolation are generally considered to have lower welfare compared to animals with access to conspecifics, and may develop undesirable behaviours such as stereotypic behaviour through lack of stimulation and restriction of natural behaviour. Therefore I designed a second experiment using the same subjects to see if the welfare of singly-kennelled dogs could be improved using a CE device known as a 'DogBox'. I measured welfare based on activity budget, stereotypic behaviour and 'optimism' before, during and after implementation of the DogBox.

During the DogBox phase, subjects foraged significantly more, and were significantly more optimistic in their interpretation of ambiguous stimuli than in baseline phases; there was a trend towards reduced stereotypic behaviour; they were observed using the device consistently over time and learned the association between touching the correct stimulus on the touchscreen and receiving a reward; and there was no evidence of habituation to the device. CE presented via the DogBox therefore seems to have had beneficial effects on the welfare of this group of kennelled dogs.

Keywords: Operant training, animal welfare, stereotypic behaviour, optimism, learning ability, dogs, Canis familiaris, captivity, isolation, environmental enrichment, cognitive enrichment, cognitive challenge, cognitive bias, DogBox.

6.2 Introduction

6.2.1 Dogs in Isolation

Dogs are a useful species to study in terms of the effects of isolation on welfare, as they are an intelligent, social, domesticated species that often spend long periods of time alone, whether in laboratories, veterinary surgeries, boarding or rescue kennels, or for police and military purposes. It is not just service or unwanted dogs either – family dogs also spend large amounts of time alone in domestic houses while their owners are out at work, which leads to the development of a number of problem behaviours associated with boredom and separation anxiety such as excessive barking and destructive behaviour, leading in turn to increasing numbers of dogs being relinquished to shelters, further compounding the problem.

The RSPCA and The Blue Cross are two large UK-wide animal welfare charities that take in, kennel and re-home unwanted dogs. In 2012, the RSPCA re-homed 12,711 dogs, while the Blue Cross re-homed 2,497 dogs. This is evidently a large number of animals experiencing kennel life each year, and there are many more shelters that re-home dogs in the UK and the rest of the world. It is estimated that each dog costs approximately £15 per day to house at a shelter including food, staffing and healthcare. One of the main goals common to all shelters is to re-home the animals quickly and successfully. However, Wells and Hepper (2000) found that sixty-eight percent of dogs obtained from a shelter exhibited some kind of behaviour problem, and of the dogs returned to the shelter (around six percent), undesirable behaviour was the reason for returning a dog in eighty-nine percent of cases. The authors suggest the implementation of behaviour therapy and kennel enrichment in shelters as ways to combat these behaviour problems developing or getting worse, in order to achieve more successful rehoming rates.

In their review of indicators of welfare of kennelled dogs, Stephen and Ledger (2006) outline many studies that point to poor welfare and evidence of stress such as increased cortisol corresponding with other behavioural indicators. The reviews by Wells (2004) and Taylor and Mills (2007) found that dogs housed in social isolation showed signs of decreased welfare such as inactivity, stereotypic behaviour (for example repetitive pacing, circling and wall-bouncing) and high levels of barking, as well as increased cortisol, an indicator

of stress. Coppola *et al.* (2006) concluded from their work measuring cortisol levels in shelter dogs that social isolation is the most stressful part of kennel life for dogs, a finding supported by previous literature (Hubrecht, 1993; Hubrecht, 1995; Coppinger and Zuccotti, 1999). Hubrecht *et al.* (1992) found that although human contact has been found to be very beneficial to isolated dogs, kennelled dogs typically receive human contact less than 3% of the time.

The length of time dogs experience the social isolation and spatial restriction of a kennel is also thought to have an effect on welfare. Titulaer *et al.* (2013) examined the differences in welfare and behaviour of short-term and long-term kennelled dogs housed at one of six shelters in the UK, and found that long-term subjects were less likely to engage in affiliative social behaviour with either humans or other dogs, and were more likely to rest more and play with toys in their kennels than short-stay subjects. Deficits in normal behaviour exhibited by dogs that experience time in a kennel such as reduced learning ability, reduced affiliative behaviour and the development of stereotypic behaviour (Titulaer *et al.*, 2013) may either be inherent to the dog and its individual life history, or it could be to do with the social isolation that being in a kennel inflicts. Titulaer *et al.* (2013) did not however find any differences in optimism (see Bateson and Matheson, 2007) between the two groups. They suggest several reasons for this. Firstly that length of stay may not be as important as experience of stay for dogs. Secondly that the decision to label seven days to three months as short-term and over six months as long-term may miss some early changes associated with a stay at a shelter, such as fluctuating cortisol levels in the first week, and the development of stereotypic behaviour patterns. Thirdly, as the method of evaluating optimism in subjects involved taking the dog out of its kennel and handling it, it was not a suitable method for either aggressive or anxious dogs.

Viggiano *et al.* (2009) found that in dogs, the social isolation resulting from being housed in a shelter environment leads to impaired cognitive abilities on a human-guided food choice task, and Udell *et al.* (2010) found that shelter dogs were outperformed by pet dogs in a human-guided object-choice task, suggesting that there is a link between the shelter environment and deficits in human-dog social interaction abilities. However, causality is difficult to pinpoint as dogs may be more likely to be in a shelter because they have inferior social and cognitive abilities. Hausberger *et al.* (2007) found that with horses, a lower

cognitive ability is linked with stereotypic behaviour, although again, causality is difficult to extract.

6.2.2 Can Enrichment Improve Kennelled Dog Welfare?

There are five main types of enrichment available to captive animals, and there can also be considerable overlap: Physical (i.e. toys, substrate materials, structures), Social (contact with humans or conspecifics), Sensory (music, scents, textures), Nutritional (type and presentation of food) and Cognitive (e.g. puzzle games and operant training).

While social housing for confined dogs is generally regarded as the best solution to the difficulties posed by the need for social contact, it is not always possible due to increased risks of injury or disease transmission (Wells, 2004). Mertens and Unshelm (1996) argued the case for dogs to be housed in groups in animal shelters as they found in a comprehensive study of 211 dogs that those in group housing coped better with shelter life than those in individual housing. While 91% of social confrontations by group-housed dogs in their study were solved by posturing, suggesting that the main aversion to group housing (aggression) is a lower risk than commonly thought, 31% of the individually housed dogs developed behavioural problems, of which 10% showed stereotypic behaviours. None of the group-housed dogs developed stereotypic behaviours in their study.

Physical enrichment such as toys (balls, squeaky chews etc.) in kennels are an easy-to-implement enrichment commonly thought to alleviate boredom arising from social isolation. However, Wells (2004) concluded that toys made little difference to the welfare of adult dogs in kennels, and suggested habituation as the reason for their ineffectiveness, especially if they are not interactive and they do not provide food rewards.

Taylor and Mills (2007) reviewed the use of visual stimulation such as televisions as tested in a study by Graham *et al.* (2005b), but found the dogs habituated quickly to this too. Ogura (2012) found that video enrichment increased the welfare of Japanese macaques by reducing abnormal behaviours, but only when they could control what they watched and when.

Stephen and Ledger (2006) and Taylor and Mills (2007) argued that predictability and control are important considerations for dogs in shelters, that may help to alleviate the stress of a kennel environment that develop

behaviours which are unhelpful for the animal and for its re-homing such as stereotypic behaviour and apathy. Predictability and control are both features of operant training, which may therefore be of benefit to kennelled dogs in terms of reducing stereotypic behaviour and other indicators of poor welfare.

Cognitive enrichment (CE) focuses upon an animal's ability to learn and find ways to adapt to its environment through appropriate cognitive challenges, providing opportunities to control and predict the outcome. Many studies have found measureable benefits from the provision of CE, such as reduced stereotypic and abnormal behaviours, and increased positive behaviours such as foraging and affiliative behaviour, and improved physical condition, in many different types of captive animals (Meehan and Mench, 2007; Manteuffel *et al.* 2009; Clark, 2011; Franks, 2012; experiment one and two in this thesis).

CE has also been found to be beneficial in terms of increasing the optimism an animal feels when interpreting stimuli in its environment (experiments one and two, this thesis). A study by Mendl *et al.* (2010) indicated a link between separation anxiety and pessimism in shelter dogs, which can be understood more clearly when thinking of separation anxiety as a negative interpretation of events – the dog displays anxiety because it might be left alone for the foreseeable future. A dog that interprets being alone in a more positive way, for example when it experiences an environmental enrichment that allows it to engage in a cognitive task, may cope better in a shelter environment, and in turn may be easier to successfully re-home. As Palestrini *et al.* (2010) point out in their study of separation anxiety in dogs, the behaviours associated with separation anxiety are strongly associated with relinquishing dogs to shelters.

Even severe stereotypic behaviour in captive animals can be greatly reduced by CE (experiments one and two in this thesis), and it may therefore be useful in preventing stereotypic behaviour from gaining a hold over shelter dogs, where social contact and other distractions are limited, often leading to repetitive pacing, barking, howling and self-directed behaviours such as excessive autogrooming (experiment three, this thesis).

6.2.3 Cognitive Bias Measurement and the Advance Key Procedure

In this set of experiments, an important change was made to the cognitive bias task used to measure welfare in experiment one and two, setting it apart from all other cognitive bias tasks to date (e.g. Harding *et al.*, 2004;

Bateson and Matheson, 2007; Matheson *et al.*, 2008; Brilot *et al.*, 2010; Brydges *et al.*, 2010; Bateson *et al.*, 2011; Rygula *et al.*, 2012; Anderson *et al.*, 2012; Anderson *et al.*, 2013 and Titulaer *et al.*, 2013). Because in experiment one and two I found that subjects learnt the go/no-go procedure relatively slowly, and suggested that the reason for this was that the S- was not aversive enough to prevent subjects from flipping all of the lids, and an active choice procedure (for examples see Matheson *et al.*, 2008; Anderson *et al.*, 2012; Anderson *et al.*, 2013) does not allow for the fact that subjects may choose to forego a reward rather than risk a punishment, I decided to implement a procedure known in the operant training literature as the 'advance key procedure' (Leyland and Honig, 1975). In addition to presenting both S+ and S- simultaneously (go left/go right) so there is always a correct answer, after subjects learnt the basic S+/S- discrimination an advance key (AK) was implemented whereby a neutral second choice was always offered: AK is in effect a choice to neutrally advance the trial to the next one without incurring any costs. S+ paired with AK should always result in S+ being chosen while S- paired with AK should always result in AK being chosen. This is very helpful when testing subject's reactions to untrained stimuli such as ambiguous probes – if they choose AK in the presence of ambiguous stimuli they are interpreting the ambiguous option as S-, while choosing the ambiguous stimuli means they are interpreting it as S+. This prevents any motivational confusion leading to over-selection (excitement, hunger, boredom, lack of exercise, lack of understanding) or under-selection (depression, anhedonia, apathy, lack of appetite, lack of understanding) for which the study by Bateson and Matheson (2007) was criticised (Matheson *et al.*, 2008), as well as risk aversion due to the possibility of punishment, for which active choice studies such as Matheson *et al.* (2008); Enkel *et al.* (2010); Anderson *et al.* (2012) and Anderson *et al.* (2013) can be criticised.

6.2.4 Aims and Hypotheses

Experiment Three

I measured the welfare of kennelled dogs at two shelters in the UK: The RSPCA, Exeter and The Blue Cross, Tiverton. The aims of experiment three

were to see if the welfare of kennelled dogs was correlated with their length of stay in kennels, and to train subjects with an operant task in order to investigate the link between stereotypic behaviour and learning ability. I predicted that length of stay would be negatively correlated with subjects' welfare, and that stereotypic behaviour would be negatively correlated with learning ability.

Experiment Four

The aim of experiment four was to see if CE would be effective in improving the welfare of kennelled dogs, as it has already been found that the welfare of group-housed animals such as pigeons improves from being given access to group- and individually-presented CE (Experiment one and two, this thesis). The DogBox described in chapter three of this thesis provides operant training to individual animals in their home environment without the need for a human trainer, reducing the input needed from staff at shelters. In order to assess the usefulness of the DogBox as a welfare improvement device, measures of welfare taken were activity budget changes including stereotypic behaviours, and optimism (Bateson and Matheson, 2007) when presented with ambiguous stimuli, across three conditions, baseline one (measures taken in experiment three), the CE condition, and baseline two, to see if behaviour changes returned to normal after CE implementation. I predicted that stereotypic behaviour would reduce, and optimism would increase during the CE condition compared with baseline conditions.

6.3 Method

6.3.1 Design (Expt. 3)

Experiment three took place between November 2010-September 2011 and took 279 days to complete, due to staggered access to dogs at both shelters and concurrently running experiment four. The experiment had two phases: the first phase comprised behavioural observations, and the second phase involved training on an operant task. The dependent variables measured were activity budget and learning ability.

The experiment followed this timescale:

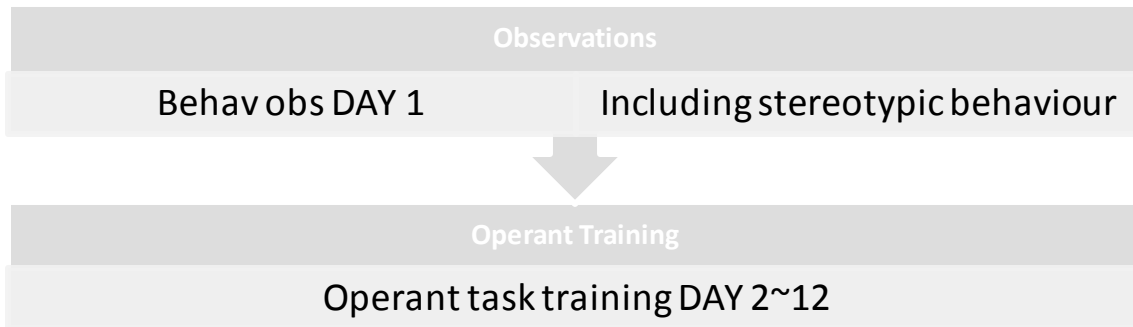


Fig. 1: Observational design of experiment three with training element.

6.3.2 Design (Expt. 4)

Experiment four took place between November 2010-September 2011 and took 279 days to complete, due to staggered access to dogs at both shelters and concurrently running experiment three. The experiment had a repeated measures, longitudinal design. Dependent variables measured were activity budget (see experiment three for full details), optimism (Bateson and Matheson, 2007) and engagement with enrichment. Fig. 2 shows the repeated measures schedule. I took the behavioural observation data collected during experiment three as baseline one. The dogs that were trained on the operant discrimination task in experiment three were then trained on a further element of the task and their optimism was assessed during each condition of the experiment. Subjects experienced five days of CE provided via the DogBox. Finally optimism was reassessed and behavioural observations were repeated for all subjects still at the shelter facilities to give baseline two.

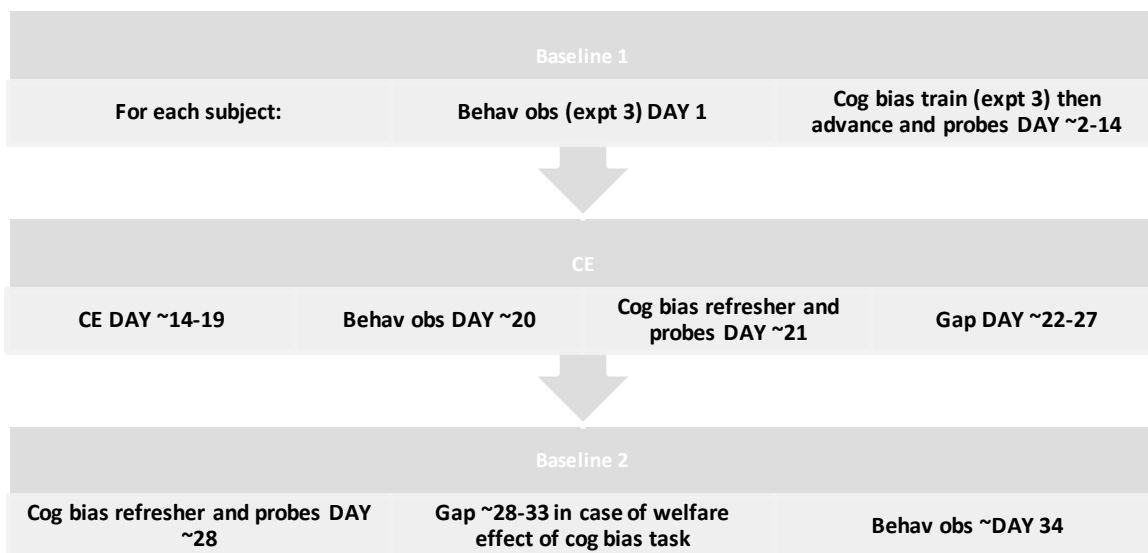


Fig. 2: The experimental schedule for each subject. Average schedules denoted by ‘~’ as each dog learned the operant task at a different rate in baseline one, therefore the schedule was staggered for each subject.

6.3.3 Housing

The study was conducted across two facilities between December 2010 and September 2011. The first facility was the RSPCA Little Valley Animal Shelter in Exeter, UK, which opened in 1991 and has facilities for around thirty-two unwanted and cruelty-case dogs, and the second was The Blue Cross Rescue Shelter in Tiverton, UK, which opened in 1985 and has a kennel block which also houses up to thirty-two dogs. The dog kennels at both facilities were similar in size and composition – they both had an indoor section constructed of painted brick walls and concrete floor with thick wire mesh doors at the front, and were approximately 7ft long x 5ft wide x 9ft high. Very large dogs had correspondingly larger kennels. The inside areas were heated in cold weather, and were each provided with a plastic bed and blankets, a water bowl and a chew-toy. In all kennels, there was a closable hatch leading to an individual outside area of similar size to the inside kennel. These areas did not have any view of other dogs. At the front, the RSPCA kennels were positioned in a row along a corridor so no visual contact with other dogs was possible there either. At the Blue Cross facility, the kennels were positioned across from each other so the dogs could have visual contact with conspecifics. However, in practice, these kennel doors often had blankets positioned over them to prevent visual contact as the staff said it caused increased anxiety in some dogs. Recent research suggested classical music can be beneficial to shelter dogs (see

Wells, 2009 for a review), so a radio set to Classic FM played softly in both facilities. Staff or volunteers took each dog out for at least one walk per day, including some off-lead exercise time if appropriate, lasting at least ten minutes. If more staff were available, the dogs got a second walk. Sociable dogs were sometimes walked in small groups. Visitors were not allowed in the kennel area at either facility, so dogs were taken to a meeting place for prospective owners to interact with them. Food was provided in a bowl in the morning and early evening in both facilities – the quantity and type was determined by each dog's needs. Water was provided *ad libitum*.

6.3.4 Subjects

Fifty-eight singly housed dogs (thirty-four male and twenty-four female) were involved in the study in total; seven at the RSPCA and fifty-one at the Blue Cross. All fifty-eight dogs were observed for phase one of the experiment and of those, thirty-nine participated in the operant discrimination task, and eleven participated in the CE condition of experiment four. The reasons for the drop-outs included re-homing, non-participation in training and euthanasia. Their ages were usually estimated by the admitting veterinarian; the mean estimated age was three years. There was a wide range of breeds of dog at both shelters including German Shepherds, Greyhounds, Collies, Labradors and Staffordshire Bull Terriers, as well as various cross-breeds. Dogs were routinely neutered at both shelters. The dogs studied at the RSPCA facility were all 'cruelty case' dogs, meaning that they had been removed from abusive conditions and were being held at the facility awaiting their owner's criminal trial. As a result these dogs had been at the facility a long time and had a history of very bad welfare. The dogs studied at the Blue Cross facility had more variable reasons for needing re-homing, ranging from elderly owners dying to strays being shipped from Ireland. We had autonomy in choosing dogs to study at the Blue Cross, so we chose dogs of various breeds, ages and backgrounds. At the RSPCA we were given access to selected dogs only as some were deemed too aggressive or anxious to study. One of the study dogs at the Blue Cross was euthanized during the study for severely aggressive behaviour to another dog. All of the others were eventually re-homed. While this is not typical, the high re-homing rate of study dogs may have been due to the fact that we chose dogs that seemed willing and interested in participating, and were not on isolation

housing blocks. The mean length of stay from entry to rehoming for dogs at the Blue Cross during the study was 16 days, while the mean stay for cruelty-case dogs at the RSPCA was 156 days. Dogs were categorised into 'Short Stay' (the median of six days or less) and 'Long Stay' (seven days or longer) based on the amount of time they had spent at the shelter when observations began. This differs from the categorisation of 'short-term' as between one week and three months and 'long-term' as six months or more by Titulaer *et al.* (2013) because, as they acknowledged, a shorter timescale may be more sensitive in picking up the early stress caused by entry to the kennels and associated behaviour changes.

6.3.5 Reliability Assessment

Observations, training and other protocols during experiment three and four described herein were conducted by myself (L. Millar) and one assistant (J. Gilbert). Inter-observer reliability was assessed by filming one hour of dog behaviour with no experimental interventions, then both of us recording behaviour on a focal-animal, instantaneous point sampling basis every five minutes. Intra-observer reliability was assessed by observing the recording three weeks later to see if I recorded the same behaviours. Pearson's correlations were conducted using frequencies of behaviour for intra-observer ($r=0.999$, $p<0.001$) and inter-observer ($r=0.999$, $p<0.001$) reliability and found that although there were two instances of disagreement there was a high degree of intra- and inter-observer reliability.

6.3.6 Activity Budget Observations (Expt. 3 and 4)

On day one for each subject we (L. Millar and J. Gilbert) carried out behavioural observations whilst sitting quietly outside the individual kennels, using the dividing walls as visual barriers and not engaging in direct eye contact so that we did not appear threatening. Each dog taking part in the study was observed between the hours of 11am and 4.30pm (access times set by the shelter facilities), alone in its kennel, building up a complete picture of activity levels including stereotypic behaviour and time spent outside the kennel. We did not interact with the dogs in any way during the observation session. Five minutes were allowed at the beginning of each observation session for the dogs to habituate to us. We recorded the behaviour being performed by the dog

every five minutes on a data sheet as measured on a timer. Behaviour categories were inactive, locomoting, alert, self-directed-behaviour e.g. autogrooming, foraging, object play, social, vocalising, stereotypic behaviour and out of kennel.

Stereotypic behaviour was defined as ‘repetitive movements or sounds such as pacing a route or barking with no obvious cause or benefit for more than thirty seconds’.

For experiment four, Observations were repeated during the CE phase, except for when CE was in place in the enclosure, in order to avoid ‘time-filling’ or observing the behaviour of subjects when engaged with experimental manipulations, which leaves less time for less desirable behaviours such as stereotypic behaviour, as criticised by Dallaire *et al.* (2012). Further observations were made for baseline two.

6.3.7 Cognitive Bias Task (Expt. 3 and 4)

The operant task had a white/black, left/right design, counterbalanced between the subjects. It formed the first part of training for a cognitive bias task (used in experiment four) based on the paradigm by Bateson and Matheson (2007), with the following alterations: the task had a left/right, white/black design, counterbalanced between the subjects, and the two stimuli were presented simultaneously.

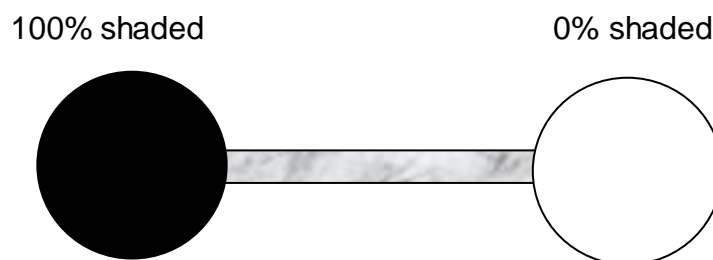


Fig. 3: Stimuli used in the operant task: a wooden stick with black and white discs attached using VelcroTM, presentation balanced equally between sides.

Stimuli (shown in Fig. 3) were two coloured discs made of thick card and protected with clear plastic laminate, 80mm in diameter. They were presented in pairs on a wooden stick pressed up against the wire mesh of the individual kennel for the dogs to touch with their noses. Half the subjects were given a small food reward (a commercial dog training treat) for touching the black

stimulus, and half the subjects were given the same reward for touching the white stimulus. The stimulus position presentation (left or right) was randomised. Experimenter effects were minimised by designing the stimuli so that the experimenter could not see which stimulus was which from the back when being held up to the bars, and only after the subject had chosen, the appropriate response (reward or time-out) was revealed, avoiding the possibility of unconsciously urging the subject one way or the other. No verbal encouragement or acknowledgment was given. The stimuli were presented for up to 20 seconds and would have been removed if not touched by then although all thirty-nine dogs in this part of the study participated fully and there were no ignored presentations. If a negative stimulus was chosen, a 'time-out' of ten seconds was observed, where a complete cessation of training took place. After the ten seconds had elapsed, the next trial continued as normal. The 30 second 'time-out' used in experiment one and two with pigeons was reduced to ten seconds for this experiment because it became evident that the dogs became very agitated while waiting. The decision to use time-out instead of a more noxious punishment such as electric shock was an ethical one, as Lea (1979) points out that it is our duty as researchers to implement methodology that causes the least amount of suffering in the subject, especially as these experiments were aiming to improve welfare.

Criterion was considered as 24 correct choices out of a possible 32 in one training session (75%). Each training session took approximately ten minutes to complete. The number of training sessions each subject required to reach criterion on the above task was recorded and taken to be their learning ability.

Advance Key Procedure (Expt. 4)

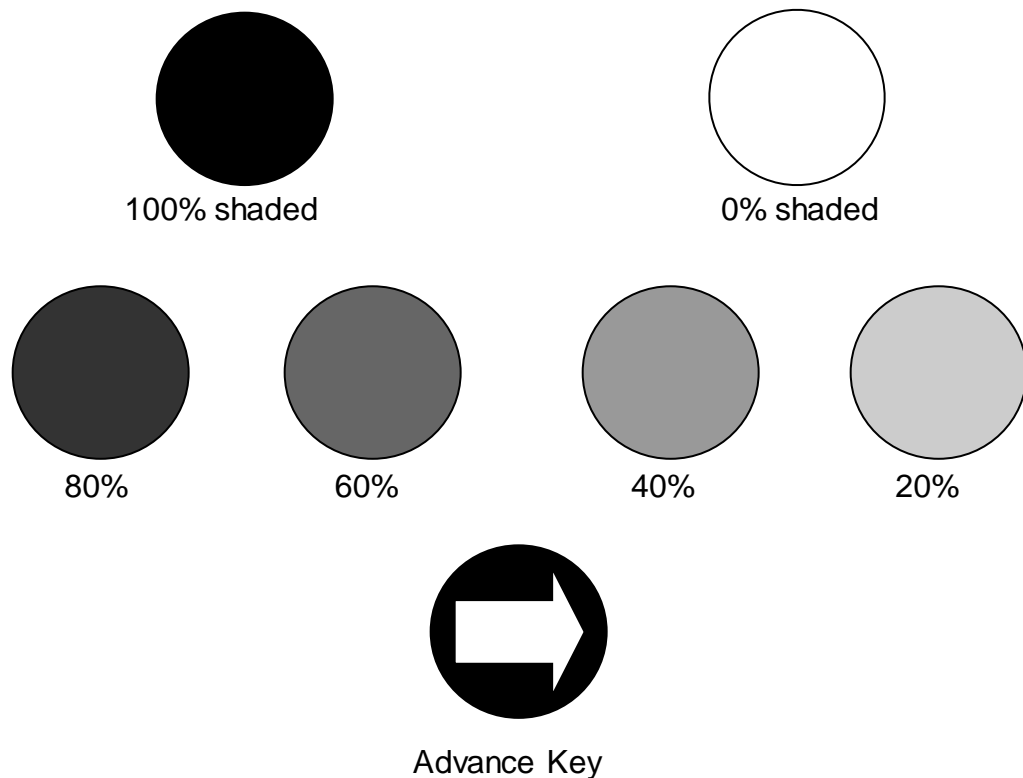


Fig. 4: Stimuli used in the cognitive bias task: the black and white S+/S- stimuli, the probe stimuli in shades of grey, and the advance key.

For experiment four, after subjects reached criterion on the S+/S- discrimination, the advance key procedure (AK), which consisted of a black disc with a white arrow (see Fig. 4), was then trained. A simultaneous presentation of either a S+ or S- and AK stimulus was used, where S+ was always rewarded, S- was always unrewarded and paired with a time-out of 10 seconds, and the AK brought about the next trial with no reward and no time out. Criterion was reached when the dog chose S+ over AK 12 times (75%), and AK over S- 12 times (75%), during one training session of 32 trials.

Probe Trials (Expt. 4)

Probe sessions were run three times in total: one during baseline one, one on the day following the CE phase, and one five days later. During each probe session, there were 32 trials – 16 probes (20%, 40%, 60%, 80%) and 16 training (eight S+ and eight S-), (see Fig. 4) each presented along with AK, in

randomised order. Each probe stimulus was shown four times and reinforced once in total during a training session as this reinforcement schedule has been found to be effective in repeated measures design experiments (see experiment one and two).

A subject touching a probe stimulus was taken as evidence that the subject interpreted the stimulus as positive, a measure of subjects' 'optimism'. Choosing the AK, where the dogs can choose to advance the trial to the next presentation, indicated a pessimistic interpretation of the current presentation. In this way, each subject was given an 'optimism score' based on their interpretation of the ambiguous probes.

6.4 Results

Data were analysed using SPSS v.19. When possible, parametric statistical tests were used, however some datasets showed significant departures from normality despite arcsine transformations, therefore non-parametric tests were implemented where necessary. Where multiple comparisons were made, Bonferroni significance corrections were used. Mdn signifies median, R is range, M is mean and SE is standard error.

6.4.1 Site, Length of Stay and Behaviour (Expt. 3)

Dogs at the RSPCA spent a significantly longer time at the shelter before taking part in the experiment than dogs at the Blue Cross (RSPCA: N=7, days at shelter M=156.14, SE=48.66; Blue Cross: N=51, days at shelter M=16.43, SE=3.97; $F(1,56)=48.20$, $p<0.001$).

The effects of two main independent variables were analysed using a multivariate ANOVA: Site (RSPCA: N=7 and Blue Cross: N=51) and length of stay at the beginning of the observations (median of six days or less: N=30, and seven or more days: N=28).

The dependent variables analysed were based on the activity budget observations and were resting, moving, alertness, autogrooming, foraging, object play, social, vocalising, stereotypic behaviour and time spent outside kennel.

There was a significant main effect of Site on object play and time out of kennel ($F(9,44)=3.86$, $p=0.001$) and of length of stay on social behaviour and

stereotypic behaviour ($F(9,44)=2.96, p=0.008$). There was no effect of either Site or length of stay upon the subjects' resting, moving, alertness or autogrooming.

The sections below expand upon these findings.

6.4.2 Effect of Site

Fig. 5 shows the amount of time dogs spent playing with toys in their kennel and time spent out of kennels in both study sites. There was a significant difference in the amount of time dogs spent playing with toys, dependent on Site (RSPCA $M=0.07, SE=0.04$; Blue Cross $M=0.01, SE=0.004$; $F(1,57)=9.35, p=0.003$), meaning that dogs at the RSPCA spent significantly more time playing with toys in their kennels than dogs at the Blue Cross.

Dogs at the Blue Cross spent significantly longer outside their kennel (walking with volunteers or interacting with visitors) than dogs at the RSPCA (RSPCA $M=0.06, SE=0.01$; Blue Cross $M=0.13, SE=0.014$; $F(1,57)=4.05, p=0.05$).

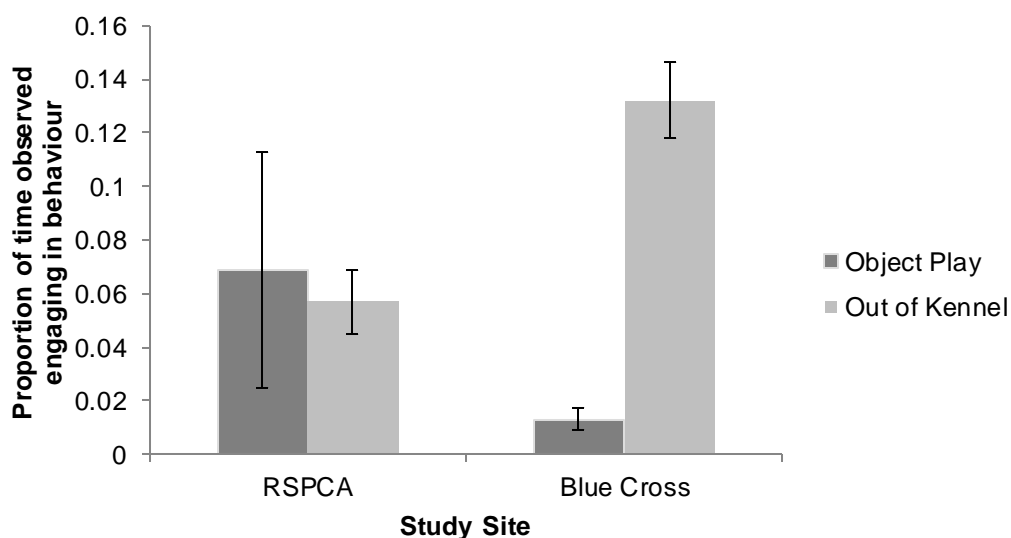


Fig. 5: Mean proportion of time spent engaging in object-play behaviour and in time out of the kennel in each study site. Error bars represent standard error.

6.4.3 Length of Stay, Social Behaviour and Stereotypic behaviour

There was a significant difference in the amount of time dogs spent engaging in social behaviour with other dogs and humans through the bars of the kennels, including play-bowing, sniffing, tail wagging and licking but not barking, dependent on length of stay (median of six or less days $M=0.14, SE=0.03$; seven or more days $M=0.05, SE=0.01$; $F(1,57)=8.90, p=0.004$). This

means that short-stay dogs engaged in significantly more social behaviour than long stay dogs. This median divide was implemented in order to maintain data normality because there were two distinct groupings in the data – short stay dogs (M=1 day at beginning of study), that were involved in the study because two large groups of dogs arrived all together, meaning that we could study them all from day one; and long-stay dogs that had typically been in the shelter for a period of several weeks (M=68 days) before the start of the study.

A final significant difference lay in the amount of stereotypic behaviour (repetitive pacing, wall-bouncing and barking) observed dependent on length of stay (at the shelter for the median of six or less days M=0.02, SE=0.007; seven or more days M=0.08, SE=0.02; $F(1,57)=7.29$, $p=0.009$). This indicates that the longer dogs stay at both shelters, the more stereotypic behaviour they engage in. Stereotypic behaviour was a common occurrence in both shelters, with the incidence in individuals ranging from 0-52% of time observed with a mean of 5% and seen in 47% of subjects in total. Stereotypic behaviour was seen in both short-and long-stay subjects, the highest level seen in the short-stay group was 8% in two dogs that had been at the shelter for only two days.

Fig. 6 shows the amount of time dogs spent engaging in social and stereotypic behaviour as a proportion of total time observed depending on whether they were characterised as short or long stay dogs.

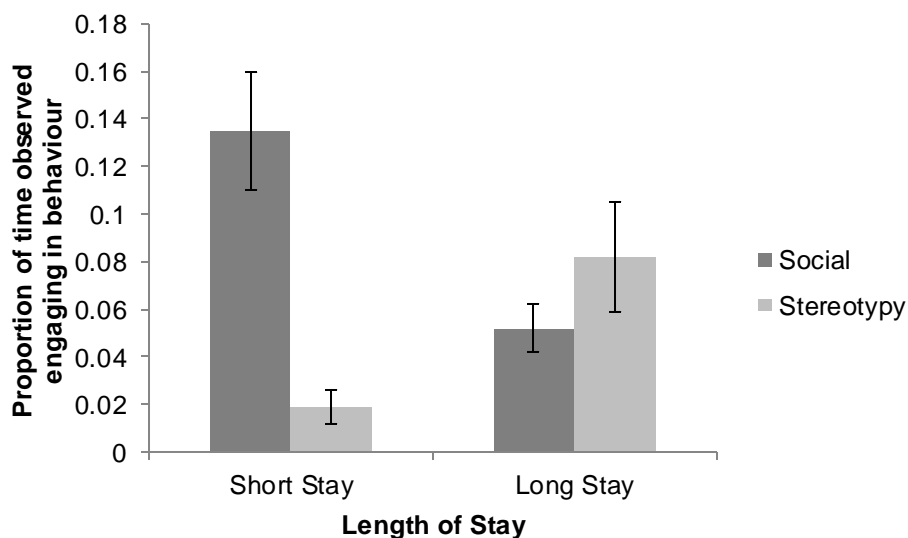


Fig. 6: Proportion of time subjects spent engaged in social behaviour and stereotypic behaviour according to whether they had been in the shelter for a short or long stay at the beginning of the experiment. Error bars represent standard error.

6.4.4 Learning Ability and Stereotypic behaviour

Thirty-nine of the fifty-eight subjects participated in the operant discrimination training task. Of the remainder, some did not engage with the task at all and were removed from the training programme, while others were re-homed before they had the opportunity to participate. The range of training sessions required for subjects to learn to criterion was 3-12 with a mean of 5.

A Pearson's Correlation conducted on the total amount of observed stereotypic behaviour and learning rate showed a significant relationship, $r=0.60$ (one-tailed), $p<0.001$ meaning that the more stereotypic behaviour a subject was observed to carry out, the more sessions were needed to learn the operant training task to criterion.

A binary split was then made in the data based on whether subjects showed any stereotypic behaviour, and a one-way ANOVA then showed a significant difference in learning rate between subjects that showed any sign of stereotypic behaviour and those who showed none. This was done because it is believed that there are differences between individuals that begin to develop stereotypic behaviour patterns, and those that do not. Subjects who showed any sign of stereotypic behaviour took significantly longer to learn the task to criterion (no stereotypic behaviour $N= 18$, sessions to criterion: $M=4$, $SE=0.40$; some stereotypic behaviour $N= 21$, sessions to criterion: $M=5.52$, $SE=2.73$; $F(1,37)=4.18$, $p=0.05$). Fig. 7 shows the number of sessions to criterion on the operant training task for dogs that showed no stereotypic behaviour versus some stereotypic behaviour.

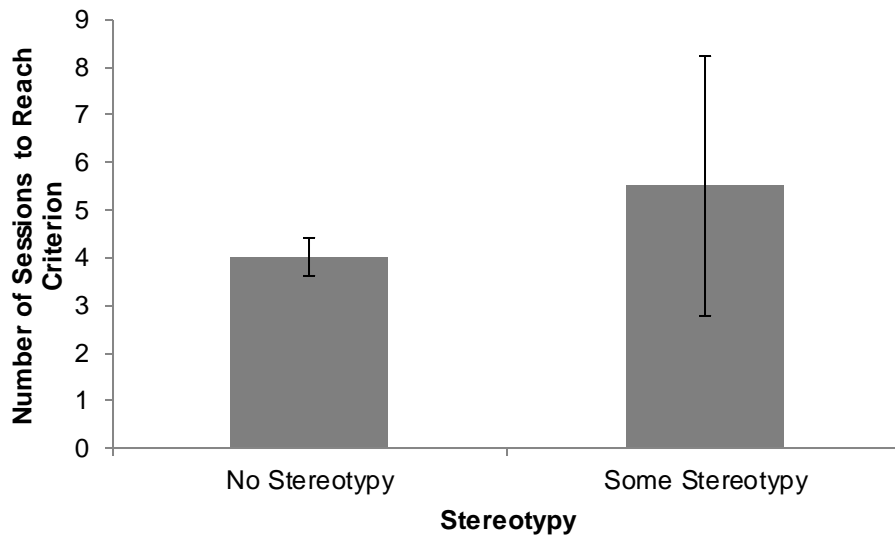


Fig. 7: Number of training sessions needed for subjects to reach criterion on an operant training task according to whether they showed any sign of stereotypic behaviour during behaviour observations. Error bars represent standard error.

A second Pearson's correlation showed a negative correlation between the amount of time subjects spent engaging in stereotypic behaviour and the amount of time spent resting, $r=-0.34$ (one-tailed), $p=0.005$, meaning that the more a subject engaged in stereotypic behaviour, the less time they spent resting.

A one-way ANOVA showed there was also a significant link between length of stay and learning ability. The longer a subject had been at the shelter, the longer they took to learn the operant training task to criterion ($F(22,38) = 4.07$, $p = 0.003$). This is represented in Fig 6. by dividing the subjects at the median of seven days stay - subjects who had been at either shelter for seven days or more at the beginning of the experiment took significantly longer to learn the operant training task to criterion than short-stay dogs (short-stay sessions to criterion: $M=3.94$, $SE=0.33$, long-stay session to criterion: $M=5.57$, $SE=0.62$; $F(1,37)=4.84$, $p=0.03$). Fig. 8 shows the number of sessions to criterion on the operant training task for short-stay and long-stay dogs.

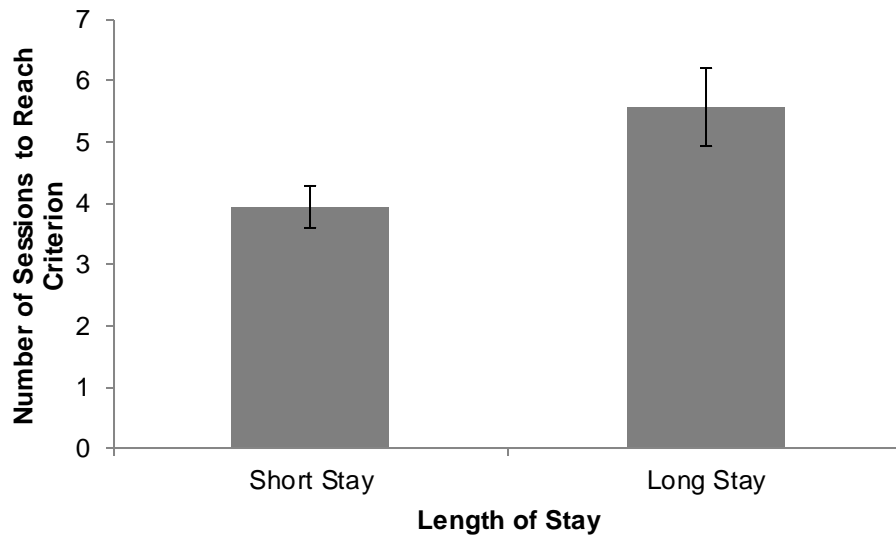


Fig. 8: Number of training sessions needed for subjects to reach criterion on an operant training task according to whether they had been at the shelter for six days or less (short stay) or seven or more days (long stay) at the beginning of the experiment. Error bars represent standard error.

As there appeared to be connections between learning ability, length of stay and stereotypic behaviour, a path analysis was conducted to further investigate the links between these elements. It indicated that learning ability and stereotypic behaviour have a strong relationship, and length of stay affected both learning ability and stereotypic behaviour development. Fig. 9 shows the strength of association of the main factors.

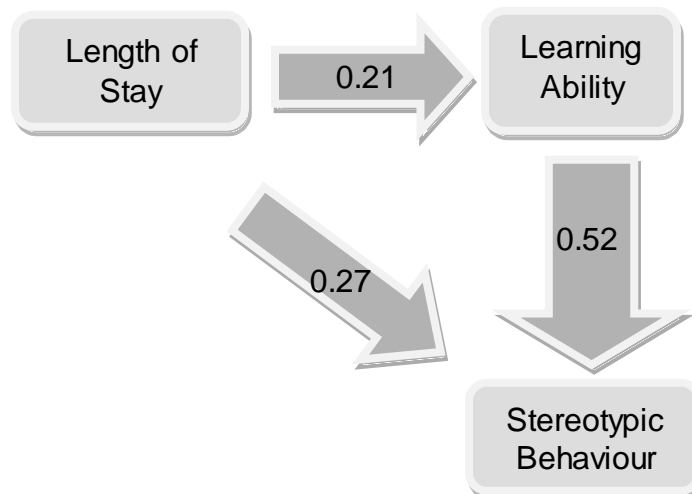


Fig. 9: Path analysis of the interactions between learning ability, stereotypic behaviour and length of stay.

6.4.5 Behaviour, Optimism and Use of Enrichment (Expt. 4)

6.4.6 Activity Budget

There were eleven subjects in baseline one and the CE condition, and in the second baseline there were only four subjects observed for the activity budget measure. Analyses were therefore restricted for activity budget for the eleven dogs involved in baseline one and the CE condition only. There were significant departures from normality despite arcsine transformations, therefore Wilcoxon tests were used, discounting the time subjects were actually engaged with the enrichment device. There was an effect of experimental condition upon the behaviour of the subjects, specifically upon foraging (Baseline one Mdn 0.02, R 0.00-0.15; CE Mdn 0.07, R 0.00-0.22; $Z = -2.37$, $p = 0.02$), fig. 10 shows the mean proportion of time subjects spent foraging during each condition. There were no significant differences between baseline and CE conditions for inactivity, locomoting, alertness, SDB's, play, social behaviour or vocalising. The analysis indicates that subjects spent significantly longer foraging (e.g. chewing edible chews and food-balls in their kennel) during the CE condition than in baseline one.

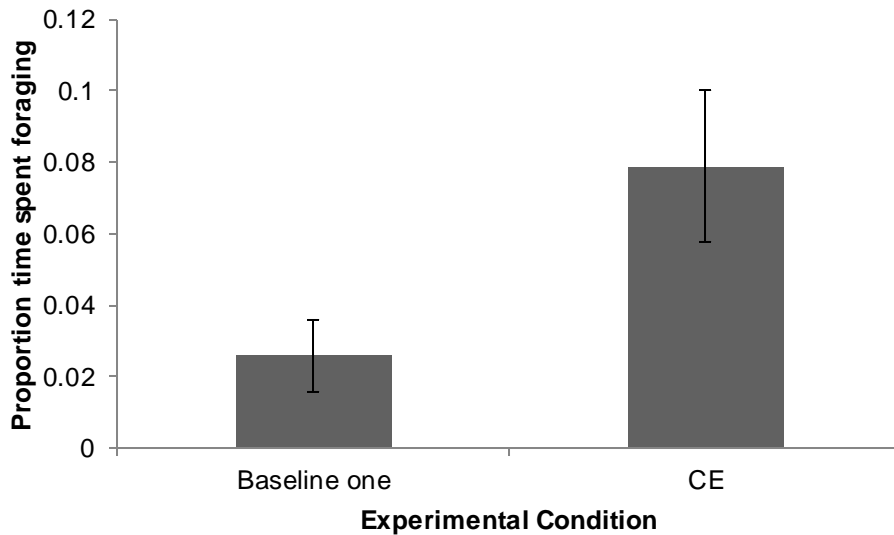


Fig. 10: Mean proportion of time subjects spent foraging in baseline one and the CE condition. Error bars represent standard error.

6.4.7 Stereotypic Behaviour

A Wilcoxon test confirmed that there was a trend approaching significance for stereotypic behaviour, ($Z = -1.83$, $p = 0.06$) as can be seen in Fig. 11, which shows the proportion of stereotypic behaviour seen in each condition (baseline one Mdn 0.00, R 0-0.52; CE Mdn 0.00, R 0-0.02). This means that stereotypic behaviour reduced between baseline one and CE measures.

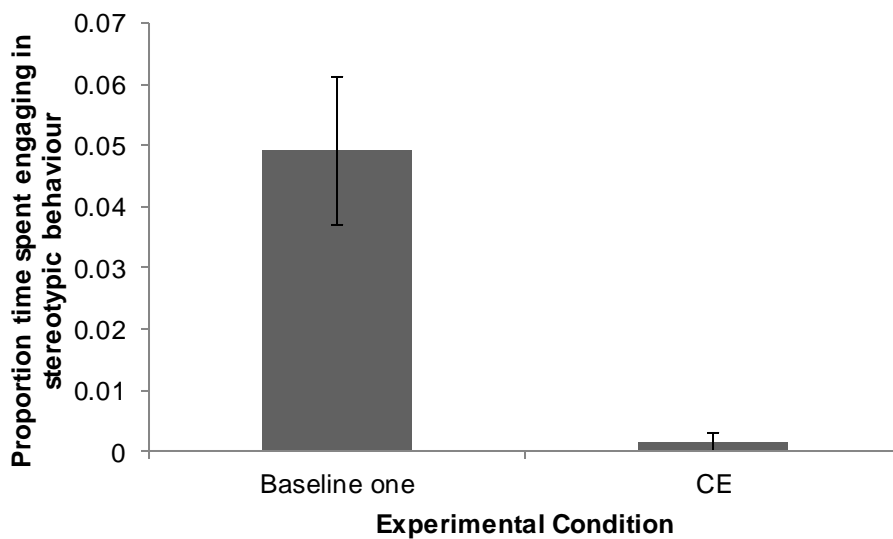


Fig 11: Mean proportion of time subjects spent in stereotypic behaviour in each condition, error bars represent standard error.

6.4.8 Optimism

All eleven subjects that took part in the CE condition of the experiment were measured for optimism before, during and after the CE condition. There was a significant effect of experimental condition on optimism scores as can be seen in Fig. 12 which shows each subject's optimism score for baseline one, the CE condition and the second baseline. A repeated measures ANOVA was conducted on the optimism scores for each condition. Mauchly's test indicated that the assumption of sphericity had not been violated ($X^2(2)=1.60$, $p=0.45$; $F(2,20)=15.78$, $p<0.001$).

A contrast revealed a significant difference between baseline one ($M=0.49$, $SE=0.04$) and the CE condition ($M=0.69$, $SE=0.04$) ($F(1,10)=30.42$, $p<0.001$). A second contrast also revealed a significant difference between the CE condition ($M=0.69$, $SE=0.04$) and the second baseline ($M=0.45$, $SE=0.05$; $F(1,10)=25.25$, $p=0.001$). A final contrast revealed a non-significant difference between baseline one ($M=0.49$, $SE=0.04$) and baseline two ($M=0.45$, $SE=0.05$) ($F(1,10)=0.72$, $p=0.42$). This means that subjects were found to be significantly more optimistic in their interpretation of ambiguous stimuli during the CE condition than in the first or second baseline.

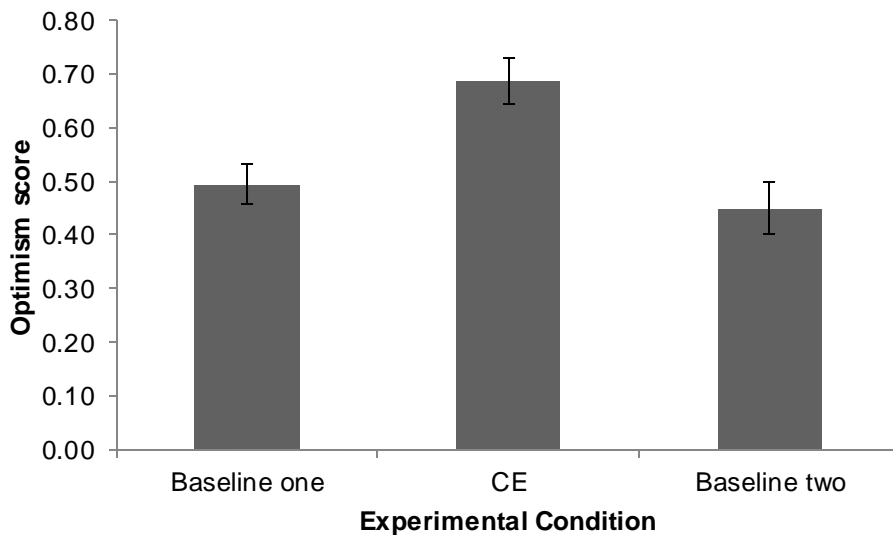


Fig. 12: Mean proportion of optimistic responses to ambiguous stimuli in each condition. Error bars represent standard error.

The optimism scores were broken down into probe gradients from 20% positive to 80% positive. Note that the experiment was counterbalanced so that

half the subjects received 0% shaded as positive whilst the other half received 100% shaded as positive – the colour palette in Fig. 13 represents all subjects' scores from closest to negative to closest to positive over the three experimental conditions. A repeated measures ANOVA was carried out on the individual probes chosen in each experimental condition. There was no interaction between experimental condition and probe choice ($F(2,30) = 1.30, p = 0.27$) meaning that the type of probe choices did not change significantly over the course of the experiment. This indicates that the reward schedule was sufficient to keep all four probe choices ambiguous over repeated presentations. If, for example, the proportion of 20% positive choices had decreased steadily over the course of the experiment, it would be evident that subjects were learning that the probe was associated with a lower chance of reward than other probes. Examination of Fig. 13 suggests that the probe closest to positive was consistently chosen by subjects at the highest rate, while the two probes closest to negative were chosen least often in all experimental conditions. A between-groups ANOVA shows that subjects' choice of the four probes differed significantly ($F(3,128) = 7.89, p < 0.0001$). Table 1, below, compares the probes with t-tests, to see where the differences lie. Significant differences were found between three combinations of probes, but not those lying closest to each other suggesting that subjects interpreted those as functionally similar.

	20	40	60	80
20(M=1.52,SE=0.21)	0	NS	t(64) = 2.69, p < 0.009	t(64) = 4.82, p < 0.0001
40(M=1.73,SE=0.26)		0	NS	t(64) = 3.77, p < 0.0004
60(M=2.45,SE=0.26)			0	NS
80(M=3.00,SE=0.20)				0

Table 1: The four probe types with t-tests, bonferroni correction set at 0.01.

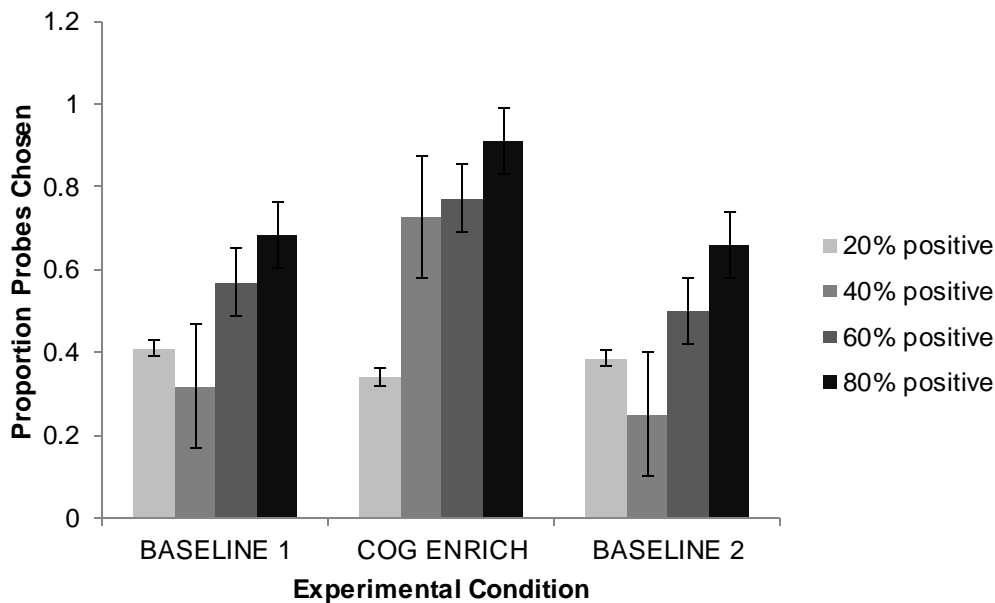


Fig. 13: Proportion of each probe gradient chosen in each experimental condition. Error bars represent standard error.

As in experiment one and two, no correlations were found between the performance of stereotypic behaviour and individual optimism scores, meaning that individual subjects displaying stereotypic behaviour were not more likely to interpret ambiguous stimuli more pessimistically than subjects that did not perform stereotypic behaviour.

6.4.9 Use of Enrichment

Subject's interaction with the DogBox increased significantly over the five presentation days, as can be seen from Fig. 14, which shows each subject's engagement with the DogBox over the twenty minutes it was available per day for five days, and summarised in Fig. 15 which shows an overall mean proportion of use by all subjects.

A repeated measures ANOVA was conducted upon the use of DogBox data collected over the five days of presentation, using a one-zero sampling method on each minute during the twenty minutes that the DogBox was available to each subject. Mauchly's test indicated that the assumption of sphericity had been violated ($X^2(9) = 25.27, p=0.003$), therefore the degrees of freedom were corrected using Huynh-Feldt estimates of sphericity. There was a significant effect of day of presentation on enrichment use ($F(2.59, 25.90)=5.86$,

$p=0.005$), which when viewed in conjunction with fig. 15 suggests that DogBox usage increased significantly over time.

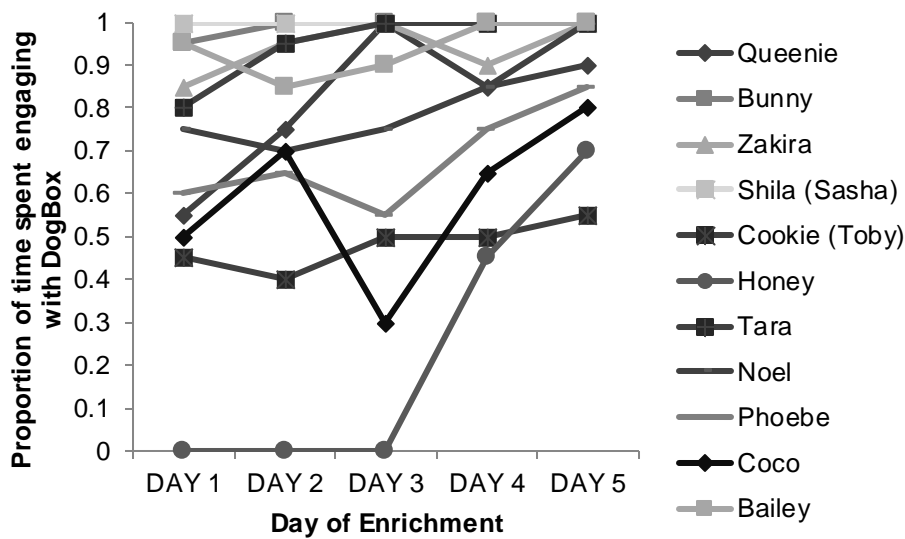


Fig. 14: Proportion of time each subject spent engaging with the DogBox over the five days of presentation, measured using one-zero sampling for each of the twenty minutes.

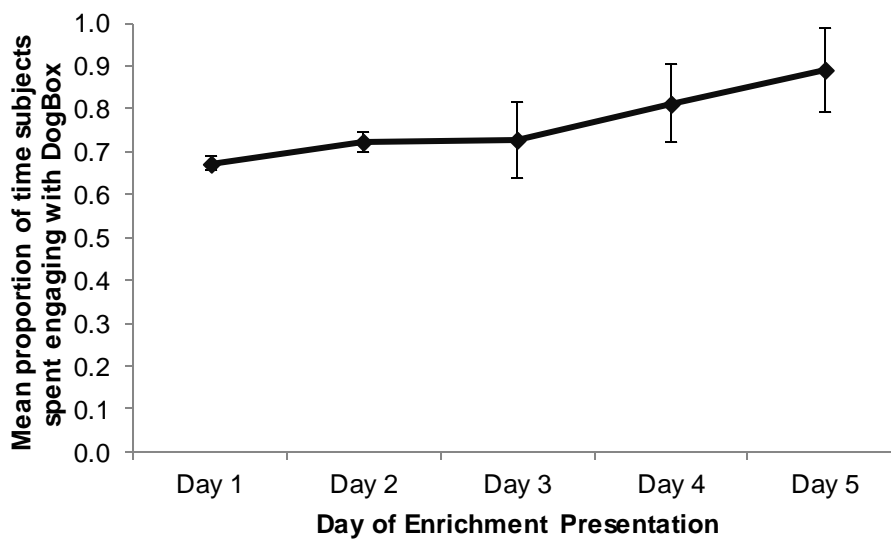


Fig. 15: The mean proportion of time subjects spent engaging with the DogBox over the twenty minutes it was available per day for five days. Error bars represent standard error.

6.4.10 Aquisition of DogBox Task

Subjects showed steady acquisition of the task presented on the DogBox screen over the course of five days' successive presentation for twenty minutes

per day, as can be seen from Fig. 16 which shows the proportion of S+ chosen by each subject on each of the five days. A maximum of 60 S+ choices were possible during each session. The S- was not paired with any outcome other than not receiving a reward therefore there was no maximum for S-. S+ selection rose steadily over the five days of presentation for nine subjects, indicating that most of the dogs were actively learning the association. Criterion, set at a proportion of 0.75 (45/60) was surpassed by two dogs out of eleven by the end of the five days, and nearly reached by a further two dogs, suggesting that the task was of adequate difficulty to hold the interest of subjects over five days but not too hard for dogs to learn. No subject moved further than stage three (see description of programme stage three in chapter three), which involved several large, static S+ presented on the screen along with several large, static S-, randomly arranged. Fig. 17 summarises task acquisition for all subjects during the CE condition.

A repeated measures ANOVA indicated that there was a significant increase in subjects choosing S+ during the course of the five days CE. Mauchly's test indicated that the assumption of sphericity had been violated ($X^2(9) = 25.27, p=0.003$) ($F(4, 1.68)=12.82, p=0.001$, Huynh-Feldt correction for sphericity). This result, in conjunction with visual interpretation of fig. 17, suggests that subjects improved their success rate over successive presentations of the DogBox.

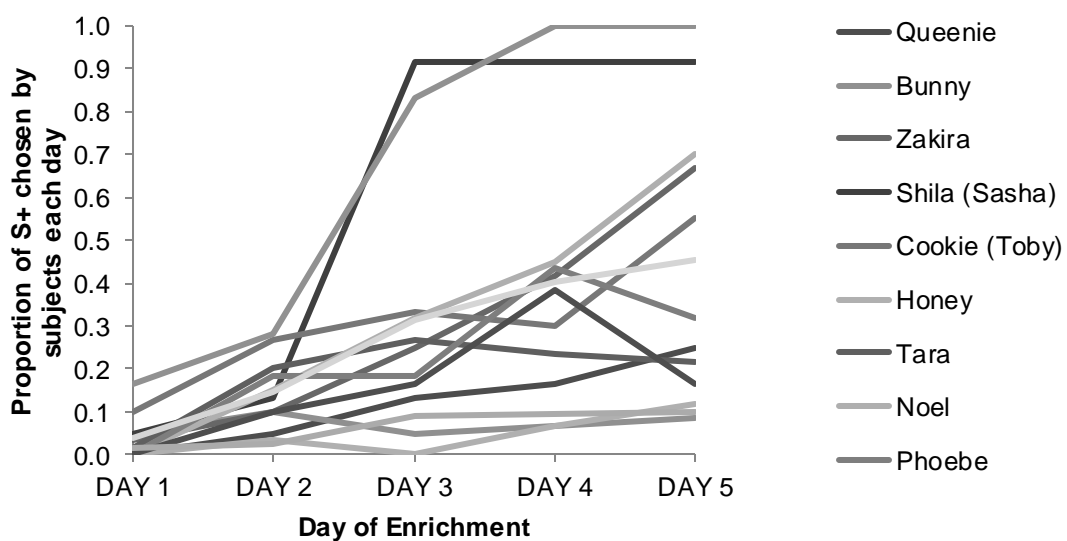


Fig. 16: Proportion of S+ touched by each subject over the five days of enrichment.

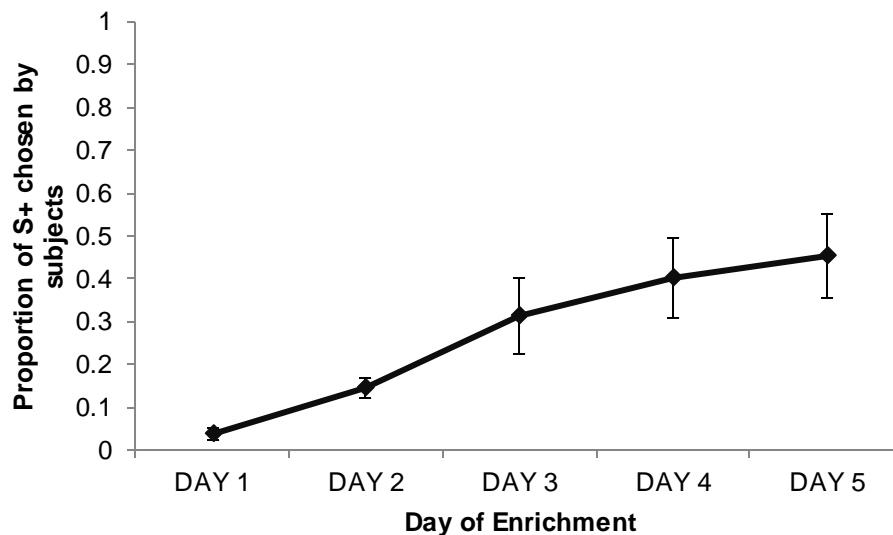


Fig. 17: Mean proportion of S+ chosen by subjects over the five days of enrichment. Error bars represent standard error.

6.5 Discussion

For experiment three, in accordance with my main hypothesis, I found that the welfare of kennelled dogs was inversely correlated with their length of stay in kennels, with a longer stay being associated with poorer welfare, although we cannot rule out the possibility that dogs at the shelter for a longer time may have inherent characteristics consistent with poor welfare. Study site also had a measureable effect on welfare, although again there were inherent differences in the dogs' backgrounds and so causality cannot be automatically assumed. I also found that stereotypic behaviour was negatively correlated with learning ability. These findings are expanded upon below.

In experiment four, the CE phase was very short; this was due to the rapid turnover of the shelter dogs, rather than being the optimum experimental duration. The study involved only eleven subjects as the rapid turnover meant that dogs were re-homed before all measures of welfare could be taken. However, despite these two shortcomings, in accordance with my aims, I found that the DogBox was effective in delivering CE to these individually-housed shelter dogs, without evidence of habituation or schedule-induced stereotypic behaviour. Dogs provided with twenty minutes per day of CE via the DogBox for five days showed a reduction in incidences of stereotypic behaviour, and an

increase in optimism and foraging behaviour compared with their baseline scores.

6.5.1 Effect of Site (Expt. 3)

The two shelters housed dogs with very different backgrounds, and on very different timescales - dogs at the RSPCA spent a significantly longer time at the shelter before taking part in the experiment than dogs at the Blue Cross. This is due to the fact that the RSPCA takes in a higher proportion of cruelty case animals – animals removed from their homes because they have been found to suffer cruelty or neglect at the hands of their owners - whereas the Blue Cross takes in mainly abandoned, homeless and stray dogs. The cruelty cases take a long time to legally process commonly leading to stays of over a year while the owners are prosecuted. After this time the dogs can usually be re-homed.

We found that the dogs' behaviour differed significantly between the two Sites: Dogs at the RSPCA spent significantly more time playing with toys in their kennels than dogs at the Blue Cross, while dogs at the Blue Cross spent significantly longer outside their kennel walking with volunteers than dogs at the RSPCA. This is thought to be due to the Blue Cross simply having more volunteers, and the RSPCA providing more toys, and rotating them more often, to compensate for this.

6.5.2 Length of Stay, Learning Ability, Social Behaviour and Stereotypic behaviour (Expt. 3)

A path analysis indicated that both length of stay and learning ability fed into the development of stereotypic behaviour. The decision to develop a path analysis model based on these factors was made because they were all found to be connected individually, and it was a reasonable assumption that a lower learning ability would incur a reduced ability to cope with captivity via less flexible coping mechanisms when faced with a longer stay in kennels. It is also possible that the propensity to develop stereotypic behaviour led to slower rehoming, or that the performance of stereotypic behaviour patterns led to a reduced ability to learn over time as behaviour became more rigid. These possibilities as well as others are expanded upon below.

The mean number of days dogs were held at the shelters was 16 days at the Blue Cross (where the majority of subjects were held), and 156 days at the RSPCA. Short-stay dogs (at the shelter for the median of six or less days at the beginning of the study) engaged in significantly more affiliative social behaviour with other dogs and humans than long-stay dogs (at the shelter for seven days or more at the beginning of the study), which is consistent with Titulaer *et al.* (2013) who found that the longer a dog is housed in a kennel, the less social behaviour it engages in and the more it plays with toys in its environment. This suggests that social withdrawal may set in with increasing time in isolation. This is not helpful to re-homing dogs successfully (Stephen and Ledger, 2006). Wells *et al.* (2002a) and Titulaer *et al.* (2013) also found that dogs that have spent a long time in kennels spend more time resting and less time engaging in social behaviours than short-stay dogs, which they agree is not helpful for re-homing. What is clearly needed is a way to get longer-stay dogs more active and less withdrawn.

A longer stay at both shelters was associated with more stereotypic behaviour such as repetitive pacing, barking and wall-bouncing. Stereotypic behaviour was a common occurrence in both shelters, with the incidence ranging from 0-52% of time observed with a mean of 5%. Even in the short-stay dogs, two individuals engaged in stereotypic behaviour for 8% of the observed time, suggesting that their previous environment had been conducive to stereotypic behaviour, as two days is probably not long enough to set up rigid behaviour patterns associated with the current environment.

Stereotypic behaviour is said to be an important measure of welfare, with higher incidence of stereotypic behaviour positively correlating with other measures of lower welfare in captive animals (Mason *et al.* 2007; experiment one and two, this thesis). These findings are in agreement with Stephen and Ledger (2006) who found in their assessment of kennelled dogs that stereotypic behaviour such as wall-bouncing and repetitive pacing, thought to be brought about by movement frustration, increased over the a time period of six weeks observation.

Stephen and Ledger (2006) tested urinary cortisol levels in shelter dogs over a thirty-one day period and found that cortisol levels peaked on day seventeen before gradually declining, while Coppola *et al.* (2006) suggest that cortisol levels rise on day two, peak on days three and four, then gradually

decline until days nine and ten, after which there is a plateau which is still significantly higher than baseline levels for pet dogs living at home. These findings indicate that the first two weeks of kennelling is the most acutely stressful time for dogs. It would therefore be appropriate to begin immediately and continue with enrichment that has a positive effect on the dogs' welfare, before stereotypic behaviour and other indicators of low welfare start to become more prevalent and entrenched.

The decision to divide length of stay by the median of six days, as opposed to the divide of one week to three months and over six months chosen by Titulaer *et al.* (2013) seemed to be effective as significant differences were found between these two groups. However, as Coppola *et al.* (2006) found through cortisol analysis that the first few days of kennelling were the most acutely stressful for dogs, it would have been expected to find more evidence of behaviours consistent with stress such as increased barking, self-directed behaviours and pacing in the short-stay dogs in this study. There could be three reasons that we did not find such behaviours in the short-stay group. Firstly there is a possibility that Coppola *et al.* (2006) did not accurately interpret the valence of the dogs' experience in shelters, and perhaps the dogs were feeling positively stimulated by the change in surroundings. This is not impossible considering that at least some dogs in shelters have arrived there from very inhospitable conditions. Secondly the human contact that Coppola *et al.* (2006) found to be helpful in reducing cortisol levels in dogs given a handling session in their first few days in the shelter may be provided in greater abundance by default at the RSPCA and the Blue Cross in this study, thus buffering the stress of the new surroundings. Thirdly, perhaps while the first few days in a kennel is indeed acutely stressful for a dog, with the change in routine, environment, social isolation and noise, it is the boredom that sets in once the dog is familiar with the shelter environment, that is more damaging to dogs in the long-term.

Length of stay was significantly linked with learning ability, with subjects experiencing a longer stay (seven days or longer at the beginning of the experiment) also showing lower learning ability, indicated by taking more time to learn a discrimination task. This is contrary to Viggiano *et al.* (2009) who found no difference between the cognitive performance of shorter- and longer-stay dogs. Instead, Viggiano *et al.* (2009) found that dogs housed individually showed cognitive deficits compared with dogs housed in groups. What needs to

be considered here are the reasons for some dogs experiencing a longer stay in kennels and indeed experiencing group- or individual-housing when both are available – causation cannot be extrapolated as lower learning ability may be correlated with behavioural problems causing relinquishment to a shelter, and/or the decision to group- or individually-house dogs within shelters.

Stereotypic behaviour was found in some of the shelter dogs but not others. Ijichi *et al.* (2013) argued that there is likely to be an individual susceptibility towards the development of stereotypic behaviour, which begins with a proactive personality type. Proactive animals have a low frustration threshold and a high motivation for rewards, and react to stress by producing more dopamine (which is involved in the development of stereotypic behaviours) than animals with a more reactive personality. This could explain why in an identical environment, some animals develop stereotypic behaviours whilst others do not.

Subjects who showed any sign of stereotypic behaviour also took significantly longer to learn the operant task to criterion. This was strongly correlated, meaning that the more stereotypic behaviour a subject was observed to carry out, the more sessions were needed to learn the operant task to criterion. Subjects also showed a negative correlation between the amount of stereotypic behaviour they engaged in and the amount of time they spent resting. These findings support the work of Hausberger *et al.* (2007), who found that stereotypic behaviour in horses was correlated with both a lower learning ability on a cognitive task and reduced time spent resting.

There may be four explanations for this. Firstly, subjects who are more likely to develop stereotypic behaviour during isolation may also be inherently slower at learning, for reasons as yet unknown. This would suggest that animals with higher learning abilities would be less likely to begin performing stereotypic behaviour in captivity. In this case, learning ability would be a good predictor of the probability of stereotypic behaviour being performed by an individual before any stereotypic behaviour is seen, and steps could be taken to prevent it from beginning. Secondly, Hausberger *et al.* (2007) suggested that reduced resting time caused by stereotypic behaviour occurring instead of resting may induce attentional problems as a cause for the lower learning ability. This is important to this study because a well-trained dog is easier to re-home, therefore the development of stereotypic behaviour may hinder the trainability of dogs and

reduce their chances of successful re-homing. This possibility would mean that a good predictor of learning ability (useful for example in laboratory studies of cognitive ability), would be the performance of stereotypic behaviour, and these animals could then be avoided. Thirdly, Mason *et al.* (2007) argue that stereotypic behaviour is independent of learning ability, and the issue instead lies with perseveration - animals that engage in stereotypic behaviour persevere with an operant task for far longer after it becomes unrewarding than non-stereotypic animals. However, this does not appear to explain the findings of either the Hausberger *et al.* (2007) study or ours, as increasing stereotypic behaviour was associated with more sessions needed to learn the operant task due to being less engaged with the procedure as well as some perseveration with the unrewarded S-. This would make testing an animal's cognitive bias difficult (see Bateson and Matheson, 2007; and experiments one, two and four for descriptions of this welfare measure), as arguably an animal that is prone to perseveration would choose probe stimuli more than an animal with a higher learning ability, thus appearing to be more optimistic than it might actually be. Perhaps a fourth possibility is that the development of stereotypical behaviour reduces the subjects' ability to concentrate on new tasks, thereby reducing their ability to learn effectively. Eilam *et al.* (2006) wrote in their review that human and non-human animals show striking similarities in the way each displays compulsive behaviours, and therefore perhaps animal stereotypic behaviour has analogous roots to human conditions such as OCD, autism and schizophrenia. Human patients with OCD frequently report an inability to concentrate on tasks due to their OCD thoughts and rituals - Nakao *et al.* (2009) found working memory deficits and other neuropsychological dysfunctions in OCD patients, and these increased with the severity of their OCD symptoms. What is not clear is if these deficits come about because of the OCD or that the deficits themselves cause the OCD.

Franks (2012) found in her study of rats that the subjects that performed better on cognitive tasks such as learning to turn lights on and off and getting rewards via a maze also showed less signs of stress such as fewer boli (faeces) produced. But, does this mean that they become less stressed because of their effectiveness at learning the tasks, or does it mean that less stressed individuals perform better on those tasks because they can concentrate on

learning better? The study failed to unpick causation but the link remains interesting.

For stereotypic behaviour to be correlated with reduced learning ability seems counterintuitive – an animal feeling so frustrated by their restrictive environment that they are driven to develop abnormal behaviours as a coping mechanism would seem to have higher intelligence than an animal content with a lower stimulation level. It would be interesting to carry out operant training with a subject sample on the first day of their stay in a kennel before any stereotypic behaviour due to kennelling has gained a hold, in order to see what their learning ability is, then study their behaviour over time and see if a higher rate of stereotypic behaviour correlates with a lower learning ability, then train them again on a different operant task and see if their learning ability has reduced. This way causality could be determined – a high learning ability dropping after stereotypic behaviour has begun would suggest that stereotypic behaviour itself reduces the ability to learn. Other measures of welfare could be taken with a large sample size to see if it is just stereotypic behaviour or general low welfare that has an effect on or link with learning ability. A low learning ability coupled with the development of stereotypic behaviour over time would suggest that a tendency towards stereotypic behaviour is linked with inherent low learning ability. This would have implications both for preventing stereotypic behaviour in kennelled dogs and in the wider field of cognitive research a case could be made that any stereotypic behaviour in subjects could bias experimental results. Subjects who had been in the shelter for seven days or more at the beginning of the experiment also took significantly longer to learn the operant training task to criterion. As there is a link with longer stay and the development of stereotypic behaviour, this may go towards explaining the phenomenon. It could be that dogs who are slower to learn may take longer to re-home: a well-trained/trainable dog is more likely to be adopted sooner. Another possibility is that the length of stay induces apathy and as a result reduces their interest in learning new tasks.

6.5.3 Activity Budget (Expt. 4)

There was no effect of experimental condition on the amount of time subjects engaged in resting, moving, alertness, SDB's, play, social behaviour or vocalizing. However, subjects spent significantly longer foraging (including

chewing edible chews and food-balls) during the CE condition than in baseline one. This did not include time spent with the device – data were collected only when the DogBox was not available. The lack of experimental effect on most behaviour patterns could be due to the small experimental sample size, the short experimental phase or lack of sensitivity of measurement. It could also be due to the fact that kennel life is very restrictive, and many behaviours performed by domestic dogs such as exploration and running can only be done when out of the kennel, therefore dramatic changes in activity levels were unlikely.

6.5.4 Stereotypic Behaviour (Expt. 4)

There was a strong trend towards reducing stereotypic behaviour between baseline one and the CE condition. This is very positive as it supports my previous findings in experiment one and two (this thesis) with pigeons, as well as indicating that CE can go some way towards alleviating the underlying stress of kennels and provide dogs with a suitably diverting activity.

However, six subjects that performed stereotypic behaviour at very high levels, up to 52% of the time observed and a mean of 27%, were also unable to learn the operant task necessary for the optimism test and were therefore excluded from experiment four. These subjects ranged from being totally uninterested in looking at the stimuli (two individuals) as stereotypic behaviour took up most of their time, to being interested in the stimuli but unable to grasp the behavioural responses necessary. The reasons for this inability to learn the task could have been because their learning ability was inherently too low, therefore predisposing them to also begin stereotypic behaviours at a high level, or that the performance of stereotypic behaviours took up too much time and concentration to concurrently learn a task, or perhaps the experience of kennelling, which had stressed them to the point of relying heavily upon the performance of stereotypic behaviours as a coping mechanism, had overwhelmed them so that they could not learn anything new.

As I found in experiment three that there was a link between stereotypic behaviour and lower learning ability, and those performing the highest levels of stereotypic behaviour did not learn the operant task at all, performing stereotypic behaviour at a high level may also have predicted non-engagement with CE which is also a kind of operant training.

In hindsight it would have been interesting to provide these subjects with CE anyway in order to see if they would have engaged with it. Although their optimism would not have been measured, their level of stereotypic behaviour could have served as a measure of welfare, and if left with the device for a longer time than the mean of ten minutes given for the operant task training, they may have become interested enough to begin learning. These dogs would arguably have been the best dogs to give CE as their welfare may have benefitted the most.

6.5.5 Optimism (Expt. 4)

Subjects were found to be significantly more optimistic in their interpretation of ambiguous stimuli during the CE condition than in baseline one or two. This pattern of optimism in subjects was also found in experiment one and two with pigeons, which suggests that CE improves the optimism, and consequently welfare, of both captive pigeons and dogs.

The fact that subjects' optimism scores returned to baseline level after the implementation of CE suggests that my method for measuring cognitive bias was effective over repeated presentations, which I also found in experiment one and two.

Our cognitive bias paradigm differed from the more usual methods (e.g. Bateson and Matheson, 2007; Titulaer *et al.*, 2013) in three ways. Firstly, as had been previously tested in experiment one and two, I rewarded the ambiguous probe stimuli on a variable interval reward schedule of one reinforced probe to every three unreinforced probes. The decision to reward the ambiguous probes in the cognitive bias task was controversial, as subjects may have learned to interpret them all positively, but I chose to do this in order to prevent the subjects from learning to interpret all the ambiguous probes as negative over repeated conditions, a criticism which was raised by Doyle *et al.* (2010) regarding the more usual cognitive bias paradigm (i.e. Bateson and Matheson, 2007) where several repetitions of unreinforced probes are presented to subjects. As this method also showed that scores returned to baseline levels when CE was removed, I argue that this reinforcement schedule is suitable for assessing subjects repeatedly when studying cognitive bias.

The second major difference in the cognitive bias measurement method in this experiment was the introduction of an advance key (AK) (Leyland and

Honig, 1975). Previous cognitive bias experiments such as Bateson and Matheson (2007) have been criticised for not accounting for other reasons subjects' choose to select an ambiguous probe or not (Matheson *et al.*, 2008), ranging from excitement, hunger, boredom, lack of exercise, to depression, anhedonia, apathy, lack of appetite and lack of understanding. These confounds may be eliminated by giving subjects an alternative neutral choice to move to the next presentation of stimuli. In the current experiment, subjects reliably chose the AK over S-, and S+ over AK, indicating that they had learned the purpose of the neutral stimulus. They also used it at similar rates in the probe sessions, suggesting that they selected AK if they interpreted a probe as S- and the probe if they interpreted it as S+. This helps in interpreting the results more clearly, and as such the AK is recommended as a helpful tool in the cognitive bias paradigm which was easy to train and did not seem to lead to any confusion.

Thirdly, a point raised by Titulaer *et al.* (2013) as a criticism of their cognitive bias paradigm used with dogs, which involved training dogs to discriminate between two spatial locations using reinforced and unreinforced bowls of food then placing bowls at intermediate points between the two, was that the method involved handling the dogs, which made it unsuitable for aggressive or anxious dogs. In addition, the upheaval of removing the dogs from their kennel and being handled and trained by unfamiliar humans in a room may have contributed to increased anxiety. My method of training for and measuring cognitive bias was far less obtrusive, as it was conducted through the mesh door of the dogs' home kennel, thus removing handling and change of environment from the paradigm. This method also involved less upheaval to shelter staff, as they were not required to unlock doors, handle the dogs or supervise, and there was no risk of danger to either the dog or the researchers. It is therefore recommended that cognitive bias training and testing with shelter dogs be done without involving relocation or handling to reduce any possible confounds as much as possible. The DogBox has the potential to be used as a non-human-led cognitive bias training device, thus further eliminating any effect of human presence on the judgement bias of subjects, as anxious dogs may feel more anxious in the presence of an unknown human trainer.

The possibility of cognitive bias training itself working as a form of CE and improving welfare whilst simultaneously measuring it also needs to be

investigated through research designed to measure the effect of cognitive bias training on welfare, perhaps by training the cognitive bias task and using other measures of welfare such as stereotypic behaviour reduction and activity budget changes to see if welfare improves solely through the application of cognitive bias training. It is reasonable to assume that as CE has been found to have a beneficial effect on the welfare of dogs, pigeons (experiment one and two, this thesis), farm animals (e.g. Manteuffel et al., 2009) and zoo animals (Clark, 2011), cognitive bias training, which involves cognitive challenge, would serve as both welfare enhancement and measure.

One criticism of the training for the cognitive bias task was that carrying out only one training session per day with each dog led to a larger-than-expected drop-out rate due to re-homing of subjects before we could implement either the CE or repeated welfare measures. In hindsight, it would have been possible to carry out two or even three sessions of thirty-two trials before the dogs became satiated or lost interest, perhaps spread out through the day, thus reducing the days needed for the dogs to reach criterion, meaning that more subjects could have experienced CE and have their welfare measured again for the second baseline.

The finding that optimism and stereotypic behaviour were not correlated with each other at all in any of the three experimental conditions is counterintuitive, as both stereotypic behaviour and optimism are measures of welfare, and they both improved overall during the application of CE and returned to baseline when the CE was removed. It counters the findings of both Brilot *et al.* (2010) who found that starlings that performed stereotypic somersaulting behaviour were more likely to interpret ambiguous stimuli pessimistically, and Pomerantz *et al.* (2012) who found that tufted capuchin monkeys that performed stereotypic head twirls were also more likely to interpret ambiguous stimuli negatively. However, Pomerantz *et al.* (2012) also found that stereotypic pacing by the same tufted capuchins was unrelated to their judgement of ambiguous stimuli, and they suggest that interpreting all types of stereotypic behaviour as having the same value as welfare indicators may not be helpful, and that perhaps stereotypic behaviour caused by environmental restrictions such as pacing due to limited space can be interpreted as necessary energy-expenditure and therefore even though it indicates lower welfare as the behaviours are unnatural and the environment

should be altered to allow normal behaviour to take the place of stereotypic behaviour, the individuals performing the behaviours may not be experiencing as poor welfare as they appear to be. Stereotypic pacing was also the most prevalent stereotypic behaviour observed in this experiment, closely followed by wall-bouncing, a similar behaviour. Perhaps the reduction in stereotypic behaviour during the CE condition indicated that they could expend their energy budget more satisfactorily using the enrichment device.

Our findings also suggest that even though some subjects did not perform any stereotypic behaviour, their cognitive bias judgements indicated improvements in their welfare due to CE analogous to the improvements made to the welfare of the individuals that performed stereotypic behaviour; therefore the absence of stereotypic behaviour in subjects should not automatically be taken as an indication of good welfare without also taking into account their optimism. This supports the use of both stereotypic behaviour and optimism as tandem welfare indicators, as individuals may respond to captivity in different ways.

6.5.6 Use of Enrichment (Expt. 4)

Both DogBox usage and successful triggers rose significantly and steadily over the five days of presentation, indicating that the subjects were actively learning the association of touching the correct stimuli for a food reward, with no sign of habituation. As the DogBox was only available for five days it is possible that usage may decline over time, this would need to be experimentally tested by providing subjects with access to the DogBox for longer periods of time. However, this may not be an issue as the programme can be made to increase in difficulty and complexity over time according to the subject's ability (see chapter three), avoiding boredom from lack of challenge or anxiety from over-complicatedness.

The allotted time for subjects to spend with the device was twenty minutes per day, with a maximum of sixty rewarded choices. This schedule had beneficial effects on welfare demonstrated by reduced stereotypic behaviour and increased optimism during the CE condition. However, it would be useful to conduct further investigations into the optimum amount of time per day the device should be available to subjects, as well as the optimum number of rewarded trials. Fagot and Paleressompoule (2009) found that baboons given

constant voluntary access to an automated testing device used it very frequently, averaging at eighty trials per bout, with some high users performing up to 1,800 trials per day. This seems an extraordinarily high number of trials, even taking into account natural foraging behaviour which takes up a large amount of time in most free-living wild animals' activity budget, and perhaps the aim of EE to provide the freedom to perform natural behaviours (UK Animal Welfare Act, 2006) is not being fulfilled by this. Indeed, welfare was not a measured outcome of the experiment by Fagot and Paleressompoule (2009) who were more concerned with whether the animals would use the device, and whether they would be able to learn the cognitive tasks.

It is unlikely that the extra food available to the dogs during training was the reason for any benefits seen, as the total amount available was at the lower end of the recommended daily treat total (www.barkerandbarkertreats.co.uk) without any adjustment of calorie intake required.

Constant access to the DogBox is unlikely to be beneficial in terms of welfare improvement as the rewards are finite, as is the subjects' capacity to consume them. This may lead to either frustration and perseveration if the rewards run out before satiation, or boredom, habituation and possible dietary problems if subjects ate just reward pellets at the expense of eating a balanced diet, and the enrichment was used at the expense of other normal behaviours such as resting.

Time out of the enclosure is considered to be the best enrichment kennelled dogs currently have, and this study found that the amount of time dogs spent outside their kennels interacting with humans and conspecifics, and exercising, was between 5.7% (RSPCA) and 13% (Blue Cross) with a mean of 12% of their observable daytime. The twenty minute slot for interaction with theDogBox was approximately 8% of the dogs' observable daytime. This suggests that twenty minutes is acceptable, although it would be useful to conduct further experiments to find the optimum amount of time for CE to be available. It is possible that a similar schedule to that of Zebunke *et al.* (2013) where pigs were called to a CE device up to thirty-three times per day for a small reward each time, would work for the kennelled dogs, as it would provide the dogs with CE throughout the day without being constantly available.

6.5.7 Conclusions (Expt. 3)

In conclusion, study site and length of stay was significantly linked with the welfare of dogs in kennels. At the RSPCA, where there were fewer staff and volunteers able to exercise the dogs, subjects spent more time playing with objects in their kennels and less time outside the kennels than dogs at the Blue Cross.

Length of stay was also significantly linked with affiliative social and stereotypic behaviour. A longer stay in kennels was correlated with reduced social interaction and increased stereotypic behaviour, both of which are unhelpful for rehoming.

Stereotypic behaviour was common overall, seen in 47% of dogs and taking up a mean of 5% of time observed. An important goal is to re-home dogs as quickly as possible to avoid these complications developing.

Dogs who showed any sign of stereotypic behaviour took significantly longer to reach criterion on the operant discrimination task. Length of stay was significantly linked with learning ability.

These findings suggest that even a fairly short kennel stay may adversely affect dogs in terms of developing unhealthy behaviour patterns, becoming less able to learn new tasks, and becoming less sociable, although it cannot be ruled out that dogs that take longer to re-home are also inherently at risk of poorer welfare, perhaps because they cannot be trained so easily due to a lower learning ability, or they already have undesirable behavioural tendencies.

6.5.8 Conclusions (Expt. 4)

Overall, the findings suggest that CE is useful for improving captive animal welfare, and the effect of CE on captive animal welfare is worthy of further investigation.

As both the RSPCA and the Blue Cross already spend a lot of money trying to recruit staff and volunteers to help exercise the dogs, an alternative (perhaps less expensive) way to reduce the development of unhealthy behaviour patterns may be to focus on the kennel environment itself and develop a method of EE that reduces the likelihood of dogs becoming less sociable and developing stereotypic behaviour patterns, which are difficult to

stop once established (Bloomsmitth *et al.*, 2007), and supports their ability to learn new tasks, perhaps by giving them operant tasks to learn without the need for human trainers.

The application of CE via the automated device known as the DogBox was effective in improving the welfare of individually-kennelled dogs at The RSPCA Exeter and The Blue Cross Tiverton UK in terms of reducing stereotypic behaviour and increasing optimism and foraging behaviour. This is a positive result and suggests that the provision of CE is an important consideration, beginning in the early days of kennelling before stereotypic behaviour and other signs of lower welfare have become entrenched, as findings from experiment three indicate this can happen very rapidly.

Subjects increased their use of the DogBox over time, learned the simple discrimination task in return for food rewards, and did not show any sign of habituation. This suggests that the DogBox was an effective means of providing dogs with CE, and should now be tested for a longer period of time, with more kennelled dogs, and also other captive animals.

The provision of CE provides dogs with an alternative coping mechanism, an engaging cognitive challenge, and a positive focus in an otherwise very unstimulating, isolating environment, and it would be interesting to see if a large-scale implementation of daily DogBox sessions for all dogs in a shelter would dramatically improve welfare for those animals.

The use of the improved cognitive bias paradigm which included the variable interval schedule of probe reinforcement enabling repeated measures, the advance key and simultaneous presentation, and finally the contact-free training method, all combined to give a robust, reliable measure that has been shown to capture the emotional states of both pigeon (experiments one and two, this thesis) and dog subjects. I therefore suggest these improvements to the cognitive bias paradigm should be utilised in future experiments aiming to assess animal affect through cognitive bias.

7.0 Chapter Five

7.1 General Discussion

The main aim of this research was to investigate the effect and value of CE upon the welfare of captive animals; and to find the most effective welfare measurement techniques in order to optimize the validity, usefulness and generalisability of the findings.

In order to achieve this I have examined the effect of CE on the welfare of captive, group-housed laboratory pigeons, and individually-kennelled dogs, in a methodical and standardized way, enabling comparisons between studies and real improvements to captive animal welfare to be made.

In following the guidelines laid down by Swaisgood and Shepherdson (2005) I used a repeated measures design where appropriate in order to assess the value of CE to individual animals compared with their own baseline welfare measurements, described the specific enrichment devices and procedures clearly including the amount of time CE was available and the amount of time subjects interacted with it, and have provided statistical information including standard error.

7.1.1 Potential Areas for Improvement

I used a fixed-order repeated-measures design schedule in experiment one with no counterbalancing which could be considered a potential confound (i.e. it would have been useful to have two groups of pigeons in experiment one that experienced baseline, PE, baseline, CE, baseline in order to compare it with the group that experienced baseline, CE, baseline, PE, baseline). However, as I took repeated baseline measures before and after any EE was provided, and those measures showed that in each case welfare returned to baseline levels when EE was not present, it seems that the order in which the EE was provided did not affect the outcome in this case. The ABACA methodology in experiment one, as well as the ABA methodology in experiment two and four allowed each subject to be its own control as there was a large range of variability between individuals in all experiments.

The potential confound in experiment two, due to subjects that were randomly assigned to the low-complexity group having inherently lower

optimism than subjects in the high-complexity group in the first baseline measure, could have been avoided by analysing the data collected in each condition at the time rather than at the end when the experiment was finished. However, although the measures differed, they did not differ significantly, and both groups showed a return to their separate baseline levels of optimism after they experienced CE, suggesting that each individual did in fact have a natural optimism level that was positively affected by CE, so actually it is quite a useful error.

A larger sample size in experiment one, two and four would have been preferable. However, only sixteen pigeons were available at the time of experiments one and two, while the dogs taking part in experiment four were rehomed so quickly that it was difficult to recruit subjects for long enough to run all three experimental conditions in experiment four.

Linked with this is the criticism that five days of CE in experiment four is not long enough to assess longitudinal welfare changes. If dogs could have been recruited for longer it would have been better to run the CE condition for at least ten days in order to mirror experiment one and two.

Also connected with the high turnover of dogs in experiments three and four is the criticism that the cognitive bias discrimination task could have been taught to the dogs more quickly if two or three sessions of thirty-two trials were run per day instead of just one. This would have enabled more dogs to experience the CE condition.

There were several dogs in experiment three that could not learn the simple discrimination task due to their high level of stereotypic behaviour performance. I disqualified these dogs from experiencing the CE condition on the grounds that if they could not learn the task, they would not benefit from the CE as they would not be able to learn the discrimination and gain rewards. In hindsight I believe this was an error as the dogs may have benefitted from slower, autonomous learning, and there is a chance that their welfare may have been improved to an even greater extent than the dogs that participated. In experiments one and two, there were three pigeons that could not learn the task required to measure their optimism. Their inability to learn corresponded to high levels of stereotypic behaviour for one subject, fearful behaviour when presented with the apparatus used in the task for another subject, and one subject that responded to every stimulus regardless of its meaning for over

thirty sessions. However I decided to include them in the EE phases on the grounds that in experiment one, the EE was presented to the group and so to remove them would have altered the dynamics of the group. The three pigeons that did not learn the task interacted with both types of enrichment device, and their level of stereotypic behaviour reduced. I therefore decided to include them in experiment two, where again they interacted with the EE and their stereotypic behaviour reduced again. In future experiments of this nature, all subjects displaying high levels of stereotypic behaviour would be included even if they did not learn the discrimination task necessary to assess optimism, as a reduction of stereotypic behaviour would be an improvement for these animals.

I controlled for observer effects and bias by conducting observations and experiments through visual barriers and by double-blind manipulations. However, as observer and experimenter effects are always possible when carrying out observation, training or testing, I would like to design a cognitive bias task to be given entirely via the DogBox for all future experiments, in order to remove the human element of training and testing. I would also ensure visual barriers are always used during observations - whilst video has its place in documenting research, I find that a lot of behavioural nuance can be lost during playback, and it is not a replacement for real-time observation.

7.2 Main Findings

This research set out to answer a set of questions presented in the general introduction, which have all been looked at in depth in the literature review and experimental sections. My answers to those key questions are set out below.

7.2.1 What is cognitive enrichment, and can it improve captive animal welfare?

Cognitive enrichment (CE) can be defined as the provision of opportunity to captive animals, in their home environment, which allows them to use their cognitive skills to solve problems, and which actually improves their welfare in some measureable way. It is distinct from other common types of environmental enrichment (EE) by the fact that it may not be physically demanding in the same way that climbing or digging may be, it does not necessarily have a social element, and it does not rely upon sensory input beyond stimulus identification.

However, there can be cross-over between the types of enrichment, for example operant training combines social enrichment in the form of time with a human trainer and CE through learning.

During this programme of research, I have conducted a series of experiments with the aim of finding out whether CE actually improves captive animal welfare. I found in experiment one that CE improved captive group-housed pigeon welfare by reducing stereotypic behaviour and autogrooming, while increasing optimism as measured using a cognitive bias task. I went on to find in experiment two that the same pigeons also benefitted from individually-presented CE, again by reducing stereotypic behaviour and increasing optimism. Finally in experiment four, I found that individually-kennelled dogs in shelters also benefitted from CE, as the dogs showed welfare improvements including reduced stereotypic behaviour and increased optimism.

My findings build upon the work of Puppe *et al.* (2007), Langbein *et al.* (2009), Manteuffel *et al.* (2009) and Zebunke *et al.* (2013) as well as the ideas posed by Meehan and Mench (2007), Clark (2011) and Franks (2012), and support the notion that CE improves the welfare of captive animals.

CE can be effectively presented using both low- and high-complexity and low-and high-tech devices (see experiment one, two and four, this thesis). The actual design of CE devices can vary hugely depending on the species and the environment, as well as the human caregiver's time, crafting ability and funds. Any CE device should be customizable in order to avoid habituation over time and to tailor it to individual skill levels to prevent frustration. A CE device should include: (a) a cognitive puzzle to solve, (b) extrinsic rewards, and (c) customizability.

7.2.2 Is cognitive enrichment more beneficial to captive animals than physical enrichment?

This question cannot be answered absolutely, as it will depend on the types of CE and PE used. However, as shown by experiment one (this thesis), a CE device was more beneficial than a PE device in the context of short-term welfare improvements for group-housed laboratory pigeons. I designed the CE device based on solving a colour discrimination to access food rewards in cups, and the PE device based on foraging through thin strips of cardboard, and tested these devices consecutively during a longitudinal study. In the CE phase,

subjects showed less auto-grooming (linked with stress), performed less stereotypic behaviour, and were more optimistic when presented with ambiguous stimuli; than in either the baseline or PE phases. Both PE and CE led to subjects being significantly more alert, and engaging in less agonistic behaviour than in baseline conditions, suggesting that both types of enrichment were similarly effective in increasing subjects' interest in their surroundings and giving them something else to do than get into fights. Both types of enrichment also reduced subjects' level of locomotion, which may have been caused by subjects conserving energy after high levels of interaction with the enrichment devices.

The reduction of stereotypic behaviour in the CE condition of experiment one was an excellent result. The reason it was so high to begin with, in a group consisting of communally-housed individuals with space to fly and perch, optimum food rations, a water bath, comfortable air temperature and so on, could be to do with their research purpose. Instead of the commonly-held viewpoint that research animals benefit from the break when not being used in cognitive experiments, it seems that animals experiencing a drop in enrichment can respond by performing higher levels of stereotypic behaviour than animals that are always housed without enrichment (Latham and Mason, 2010). As the pigeons had previously been used in several cognitive experiments and had therefore experienced an enriched environment with the opportunity to develop cognitive skills that they would not have done in an unenriched environment, their 'dormant' time before this experiment where they could no longer access the cognitive puzzles may have led to increased frustration, leading in turn to increased stereotypic behaviour. An ethical consideration both for this experiment and the use of animals in cognitive experimentation in general is the continued provision of EE if animals are kept at a research facility for the duration of more than one experiment.

Experiment one also indicated an increase in optimism in subjects when presented with CE, which further supports the conclusion that CE improves captive animal welfare in a holistic way rather than by alleviating one aspect of poor welfare (i.e. stereotypic behaviour). The cognitive bias task paradigm that I used seemed to be a very sensitive measure of changes in affect due to experimental manipulation. Subjects used the CED significantly more than the

PED, and it was this increased use that was a predictor of the increase in optimism.

From the findings of experiment one, it can be concluded that in this case, CE was indeed more beneficial than PE in terms of increasing the welfare of captive group-housed pigeons, as it appeared to prompt increased use, which in turn seemed to be the avenue by which welfare was increased. More research is needed of course in order to create a functional framework for CE implementation for other species, settings and timescales, but in terms of the outcome from this one experiment, the CE device performed as anticipated. More research would be necessary to assess the longer-term benefits of the enrichment device, and it is probable that some alterations would be needed over time, for example to increase the size of the device and number of food cups, or change the discrimination periodically to keep the subjects interested.

7.2.3 Is increasing environmental cognitive complexity associated with increasing welfare benefits?

I found that subjects in experiment one used the CE device for more time than the PE device, and an interpretative difficulty in regard to experiment one was the possibility that the CE device was more difficult to use physically and therefore took up more of the pigeons time, therefore increasing welfare not by its intended use exactly, but due to increased physical movement. Was it mind or body exercise that improved the pigeons' welfare?

Experiment two examined the effects that the complexity of the CE device presented to subjects had on the welfare of captive pigeons. The two CE devices were copies of the CE device used in experiment one and differed only in the complexity of their cognitive challenge. Experiment two indicated that again, the welfare of subjects improved during the provision of CE, but the results were rather complicated. I predicted that while similar welfare benefits as experiment one were expected in both low and high complexity CE groups, high-complexity CE would be more beneficial in terms of reducing stereotypic behaviour and increasing optimism due to the higher cognitive challenge. I found that for both low and high complexity enrichment, subjects used the devices as intended, stereotypic behaviour reduced significantly, and optimism scores for subjects increased significantly, in line with findings from experiment one. The high complexity group used their enrichment significantly more over

time however, with no evidence of habituation. Higher use of enrichment was associated with a larger improvement in optimism. The main benefit to increasing the complexity of a CE device may therefore be that higher complexity seems to increase the amount of time that subjects use it, with less habituation over time than with less complex enrichment.

7.2.4 Is enrichment better delivered in a group or individually accessed setting?

Experiment one looked at the effect of group-presented CE with pigeon subjects, which then became the subject group for experiment two, looking at the effect of individually-presented CE. It was predicted that mode of presentation of the enrichment (group versus individual) would affect welfare, but the direction of effect was not predicted: either group enrichment would be more beneficial because of increased social interaction, or individual enrichment would create an environment of safety as dominant individuals would not be able to control access to the devices. I wanted to see how the different enrichment settings affected the same individuals so I could compare the two techniques with less confounding variables.

There was a trend approaching significance for individually-presented CE reducing stereotypic behaviour more than group presented CE. However, the subjects were displaying higher stereotypic behaviour in experiment two than in experiment one to begin with, possibly due to their experience of CE and subsequent lack of it when experiment one ended causing a surge in frustration-related stereotypic behaviour.

Both group and individually presented CE increased optimism in subjects by roughly the same amount. This suggests that it is the enrichment itself rather than the group/individual setting that increases optimism in subjects.

Subjects did not begin to use the individually presented CE devices until day five, suggesting that either the isolation prevented subjects from learning how to use the device from others, lacking the social facilitation available to them in experiment one, or that the subjects were more anxious in individual cages than in a group and this put them off interacting with the device until they got used to it.

The dominance hierarchy was affected during the enrichment phase, with more agonistic encounters during the observation period in the CE

condition than in baseline conditions. This suggests that group presentation may have been preferable as it did not cause as much upheaval by getting subjects out of the aviary each day.

In conclusion, both group and individually presented CE devices reduced stereotypic behaviour and increase optimism in subjects, suggesting that it is CE rather than the setting that is beneficial, and that both group and individually housed animals should be given access to CE in their home environment. However, group presentation seems to be favourable on balance due to social facilitation promoting early use of the device and the lack of social upheaval. Therefore, if animals are in a stable group they should receive CE together in their home environment if this does not cause problems, and any CE device should at least be large enough to allow all individuals to participate simultaneously, allowing social learning, facilitation and interaction to occur which may further enrich the enrichment experience. If animals are individually-housed they could either receive CE in their individual cages, or if socialization with other conspecifics is possible a 'CE room' could be provided.

7.2.5 What affects the welfare of individually kennelled dogs?

In experiment three, I investigated the effects of kennelling on shelter dogs. The welfare of kennelled dogs was negatively correlated with their length of stay in kennels, with a longer stay being associated with poorer welfare, indicated by an increase in stereotypic behaviour and a decrease in affiliative social behaviour, although the possibility that dogs at the shelter for a longer time may have inherent characteristics consistent with poor welfare could not be eliminated.

It is widely thought that social isolation is the most damaging part of kennelling to dogs, by nature very social animals, and unfortunately in the kennel environment, human and conspecific interaction accounts for very little of the total time available.

Subjects who showed any sign of stereotypic behaviour also took significantly longer to learn an operant task to criterion. It is possible that dogs who are slower to learn may take longer to re-home as a well-trained/trainable dog is more likely to be adopted sooner. A low learning ability coupled with the development of stereotypic behaviour over time would suggest that a tendency towards stereotypic behaviour is linked with inherent low learning ability. This

would have implications both for preventing stereotypic behaviour in kennelled dogs and in the wider field of cognitive research, where a case could be made that any stereotypic behaviour in subjects would bias experimental results.

In summary, in this study, the longer-stay kennelled dogs were more socially withdrawn, developed more stereotypic behaviour and showed deficits in learning ability compared to shorter-stay dogs. A diverting activity is clearly needed to help the animals cope with the stress of isolation, other than increased social contact, as this is unfortunately very difficult due to staffing and volunteer levels at shelters. It would therefore be appropriate to begin immediately and continue with enrichment that has a positive effect on the dogs' welfare, before stereotypic behaviour and other indicators of low welfare start to become entrenched.

7.2.6 Is it possible to develop an automated cognitive enrichment device that prevents habituation and is easy to implement in many laboratory, zoo and companion animal settings (e.g. kennel, home environment) that improves (and is also capable of measuring) welfare?

The DogBox has been developed as part of this programme of study as a simple, flexible way to provide animals with CE in their home environment. It has been found to be a valuable CE device for captive individually-housed dogs (experiment four this thesis), and has been designed to be effective, adaptable, multi-purpose, accessible, useful, safe, portable and unobtrusive. The next step is to test it with a variety of animals in a variety of captive situations, and to investigate the possibility of providing cognitive bias tasks via the DogBox in order to measure welfare without the need for human expertise in training.

7.2.7 Can the welfare of kennelled dogs be improved by cognitive enrichment?

Experiment three showed that a stay in kennels was associated with welfare deficits. In particular a longer stay was linked with increased social withdrawal, increased stereotypic behaviour and lower learning ability, although there may be inherent differences in dogs that experience a longer stay in kennels.

Experiment four indicated that kennelled dogs provided with twenty minutes per day of CE via the DogBox for five days showed an increase in optimism as measured using a cognitive bias test and a reduction in incidences of stereotypic behaviour, compared with their own baseline scores.

Subjects quickly learned the association between touching the correct stimulus and receiving a food reward from the DogBox, and their use of the device increased steadily over the five days of the experiment. No evidence of habituation was seen.

This supports my previous work carried out in experiment one and two, as well as providing some important evidence to support the introduction of regular CE in dog kennels as a way of alleviating the stress of the kennel environment, reducing the development of stereotypic behaviour, and increasing optimism.

Further research is needed to investigate the effects of longer-term CE on the above parameters, as well as any possible effects on learning ability. It is expected that the provision of regular CE would stimulate learning ability, perhaps by inhibiting the development of stereotypic behaviour if this is the causal relationship.

7.2.8 Is stereotypic behaviour a good measure of welfare and can it be reduced through cognitive enrichment?

Stereotypic behaviour has been found in this series of experiments to be a useful, informative, sensitive measure for assessing welfare changes through EE. It varied reliably in each condition, and reduced significantly when subjects were presented with CE in experiment one and two and in experiment four there was a trend towards stereotypic behaviour reducing with CE. While stereotypic behaviour can sometimes be a 'welfare scar' of past experiences, this did not reduce its significance in these experiments. It is suggested that stereotypic behaviour, if present, should be reduced as a priority in order to consider EE effective.

As I learnt from observing the pigeons in experiment one and two, stereotypic behaviours tend to develop around rigid husbandry such as feeding times. Manteuffel *et al.* (2009) suggest that in order to prevent stereotypic behaviour developing in response to CE, it should be accessible at random times.

Optimism and stereotypic behaviour did not correlate in experiments one, two or four, suggesting that stereotypic behaviour may not indicate poor welfare in the sense of a negative experience for the animal, as it may serve as a way to expend energy. It is still a reliable indicator of welfare however, as an animal driven to stereotypic behaviour is being frustrated by its environment in some way.

7.2.9 Does the cognitive bias measurement paradigm suit this kind of research, what are the current weaknesses and how can they be improved upon?

Using judgement bias to probe the underlying emotional state of animals is proving to be a sensitive, useful, valid and reliable measure of welfare in many studies, and is also supported by my findings, where optimism varied reliably according to experimental condition and in accordance with other welfare measures. In experiments one two and four (this thesis), the probe stimuli were rewarded on a variable interval schedule of one reinforced to three unreinforced, and did not have a time penalty imposed if opened. This differed from most other cognitive bias studies (e.g. Bateson and Matheson, 2007), where probe stimuli were not rewarded at all. The rationale for choosing to reward the probe stimuli in my studies was that if using the same task over five phases, the subjects were likely to learn to interpret the probe stimuli as negative if they were never paired with a reward, and give a false down-trend of optimism over the phases (Brilot *et al.*, 2010; Doyle *et al.*, 2010). We had previously used this schedule of reinforcement in Wills *et al.* (2009) with pigeons and squirrels and we found that in that experiment subjects did not learn to interpret ambiguous stimuli either as all S+ or S-, which would have led to false conclusions.

Optimism scores did indeed fluctuate between conditions in all three experiments using the measure (one, two and four), suggesting that subjects did not learn a positive association, and the probes remained ambiguous. I also used the same pigeon subjects in both experiment one and two in order to compare the effects of group- and individually-presented CE, which could have been a problem when using a cognitive bias task as a repeated measures welfare test, as there were eight repeats of the probe test given to the same subjects in experiment one and two combined. I found that optimism fluctuated

reliably between baseline and experimental conditions, returning to a fairly stable baseline each time enrichment was removed. This suggests that my reinforcement schedule meant that subjects did continue to view the ambiguous stimuli as intended – ambiguous. If subjects had learned a negative association for the probe stimuli over time a more pessimistic score would have been expected for both phases of experiment two compared with experiment one. It is suggested that this reinforcement schedule is suitable for studying subjects repeatedly when studying cognitive bias.

The advance key procedure (AK) was learned readily by subjects in experiment four, that is they reached criterion by selecting AK in the presence of a S- and by selecting S+ in the presence of AK, thus establishing that it is a better alternative than S- and a worse alternative than S+, within a few training sessions. Subjects also used the option during probe trials, selecting it in the presence of some ambiguous probes but not others. The presence of an AK allowed subjects to always perform an action rather than waiting the trial out, which enabled faster learning and less frustration, and was a more ethical solution than the active choice paradigm, which did not take into account that subjects may choose to avoid the risk of punishment over gaining a reward.

I developed a non-contact method for use with kennelled dogs involving the presentation of stimuli to dogs through the mesh of their kennels, removing any safety concerns or confounds due to increased anxiety of nervous dogs which may have been a problem if animal-human contact was required by the procedure (e.g. Titulaer *et al.*, 2013). This made cognitive bias measurement safer and more reliable, as well as quicker and easier as no assistance from shelter staff was necessary.

Time-out as a punishment may take longer to elicit the correct avoidance/response behaviour than more noxious punishments such as electric shocks, but learning does still occur (e.g. Wills *et al.* 2009). I found in experiment one and three of this thesis that most pigeon and dog subjects learned the initial discrimination although it took between 6-20 sessions (pigeons) and 3-12 sessions (dogs) to reach criterion. I chose not to use electric shocks or other aversive punishments as I felt it was unnecessary.

Ethically, humans carrying out research on animals have a responsibility to inflict the least amount of pain and suffering in the pursuit of knowledge, therefore the use of techniques such as social or nutritional deprivation, electric

shocks or other noxious punishments should be used only if there is no other way to elicit the required behaviour (Lea, 1979), and indeed the Home Office guidelines for the care of animals in scientific procedures, known as the Animals (Scientific Procedures) Act 1986, was amended in January 2013 to include the specific caveat that "...the methods used in regulated procedures applied to such animals must be refined so as to eliminate or reduce to the minimum any possible pain, suffering, distress or lasting harm to those animals" (ASPA Act 1986 Amended, 2013, p. 15).

If there is any doubt that electric shocks can be considered to do harm to animals, Branch *et al.* (1977) argue that whilst time-out and electric shocks may at first examination seem to be broadly equivalent in terms of behaviour elicited, when pentobarbital (a tranquilising barbiturate) was administered to pigeons trained to peck for rewards using either time-out or electric shock as punishers, the rate of responding was increased for electric shock subjects but not for time-out subjects. This suggests that electric shocks work by eliciting fear of sudden pain in subjects, which was ameliorated by the tranquiliser whereas time-out does not. The restriction of access to reward (a time-out) is a natural consequence to incorrect behaviour which triggers learning, working as a deterrent without causing a fear response in the subject.

In summary, if ambiguous probes are going to be presented on more than one occasion, the reinforcement schedule described above should be implemented, an 'advance key' should be used in order to allow subjects to always perform an action, the non-contact method developed in experiment four should be used in order to prevent any confounds or dangers associated with animal-human contact, and the least noxious training method able to elicit the required behaviour should be used in order for welfare to be optimised.

7.3 Conclusions

1. A cognitive enrichment procedure was found to be more effective than a physical enrichment procedure in terms of increasing desirable behaviours and optimism and decreasing stereotypic behaviour in captive group-housed pigeons. The low-tech CE device I designed was fit for purpose.
2. A moderately difficult puzzle with a food reward to aid motivation was found to be appropriate CE for captive pigeons. Too easy and there is a risk of habituation, too difficult and there is a risk of frustration. This can be probably applied to most captive animals but requires further testing.
3. Cognitive enrichment was found to be beneficial to captive pigeons in both an individual and group-presented setting, although group-presentation caused less social upheaval and allowed for social facilitation, and is therefore the preferred mode at least with pigeons.
4. Kennelled dogs may be adversely affected by a kennel stay, presumably by the unchanging, unchallenging nature of captivity and by the lack of social contact. It is possible that the longer the stay, the more damage might be done in terms of reduced sociability and increased stereotypic behaviour, both detrimental to rehoming. Stereotypic behaviour was also associated with reduced learning ability which may add to the difficulty of rehoming. It is important to remember that dogs that are relinquished to shelters, and dogs that spend longer at shelters, probably have inherent differences and therefore causation cannot be determined without additional longitudinal research.
5. The welfare detriment a kennelled dog experiences can be reduced through the provision of the CE device known as the DogBox, which reduced stereotypic behaviour and increased optimism in subjects during the five days it was available.
6. The DogBox has been found to be a valuable CE device for individually-kennelled dogs, and has been designed to be effective, adaptable, versatile, empowering, accessible, useful, safe, portable and unobtrusive. The next step is to test it with a variety of animals in a variety of captive situations.

7. Stereotypic behaviour is a useful, informative measure for assessing the effect of EE including CE on the welfare of captive pigeons and dogs. It varies reliably and in the expected direction according to experimental manipulation. While stereotypic behaviour can sometimes be a 'welfare scar' of past experiences, this did not reduce its usefulness in these experiments. Stereotypic behaviour may not necessarily indicate an animals' subjective experience of poor welfare, as shown by the lack of correlation within individuals with optimism, but it is nevertheless an indication of a suboptimal environment as it is caused by the frustration of natural behavioural urges. It is suggested that stereotypic behaviour, if present, should be reduced by EE as a priority in order to consider EE effective.
8. The cognitive bias paradigm, as characterized by optimistic and pessimistic responses to ambiguous stimuli, is also a valuable, insightful measure of welfare, which seems to reliably indicate the underlying emotional state of animals. In the experiments described here, optimism varied reliably according to experimental condition. The cognitive bias paradigm developed from Bateson and Matheson (2007) has been improved in three ways: firstly if ambiguous probes are going to be presented on more than one occasion, a partial reward schedule should be used to prevent subjects learning a negative association for the ambiguous probes. Secondly, there should be an 'advance key procedure' in place so that subjects can give a true and risk-free 'no' instead of a go/no-go or traditional choice design. Finally the non-contact method of presentation used in experiment four should be used with animals that may be affected by the presence of a human or present a danger e.g. anxious or aggressive animals, large zoo animals.

7.4 Future Aims

Experiments one, two and four show that CE improves captive animal welfare, and the development of both the simple CE devices used in the pigeon experiments and the more complex DogBox show that CE can be incorporated into any captive animal enclosure, without any technical expertise in the case of the simple devices, and without having to cause the animals undue stress by moving them out of their home environment. The next experiments carried out

in this series will be aimed at developing and testing appropriate CE for a range of species and settings.

Although only pigeons and dogs were used in this research, there is no reason why the equipment and findings cannot be generalized to other captive animals as well, with some species-specific tweaking. The use of valid, repeatable, generalisable welfare measures such as stereotypic behaviour and cognitive bias seems to be a good way to assess the real-world usefulness of enrichment devices, which is what will actually benefit the captive animals that need them.

CE should not automatically be valued more than for example access to the outdoors or social experiences, and should not be thought of as a panacea for all captive animals, rather a way to make the best of sub-optimal captive environments, particularly for social animals such as dogs being kept in individual kennels.

I aim to continue research into the benefits of CE upon the welfare of a variety of captive animals to find the best way to provide the optimum care we are duty-bound to provide.

8.0 References

- Adams, K.M. (2007). Refinement in the literature: searching for environmental enrichment. *AATEX*, 14, 307-312.
- Anderson, M.H., Hardcastle, C., Mufano, M.R. & Robinson, E.S.J. (2012). Evaluation of a novel translational task for assessing emotional biases in different species. *Cognitive, Affective and Behavioral Neurosciences*, 12, 373-381.
- Anderson, M.H, Mufano, M.R. & Robinson, E.S.J. (2013). Investigating the psychopharmacology of cognitive affective bias in rats using an affective tone discrimination task. *Psychopharmacology*, 226, 601-613.
- Animal Welfare Act 2006 <http://www.legislation.gov.uk/ukpga/2006/45>
- Asher, L., Davies, G.T.O., Bertenshaw, C.E., Cox, M.A.A. & Bateson, M. (2009). The effects of cage volume and cage shape on the condition and behaviour of captive European starlings (*Sturnus vulgaris*). *Applied Animal Behaviour Science*, 116, 286–294.
- Azevedo, C. Ciprets, R & Young, R. (2007). Environmental enrichment: a GAP analysis. *Applied Animal Behaviour Science*, 102, 329-343.
- Barrett, L.F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science*, 1, 28–58.
- Barrett, L.F., Lindquist, K.A., Bliss-Moreau, E., Duncan, S., Gendron, M., Mize, J. & Brennan, L. (2007). Of Mice and Men: Natural Kinds of Emotions in the Mammalian Brain? A Response to Panksepp and Izard. *Perspectives on Psychological Science*, 2(3), 297–311.
- Basile, B.M. & Hampton, R.R. (2011). Monkeys Recall and Reproduce Simple Shapes from Memory. *Current Biology*, 21(9), 774-778
- Bateson, M. & Matheson, S.M. (2007). Performance on a categorisation task suggests that removal of environmental enrichment induces 'pessimism' in captive European starlings (*Sturnus vulgaris*). *Animal Welfare*, 16, 33-36.

- Bateson, M., Desire, S., Gartside, S.E. & Wright, G.A. (2011). Agitated Honeybees Exhibit Pessimistic Cognitive Biases. *Current Biology*, 21, 1070–1073.
- Baumans, V. (2005). Science-based assessment of animal welfare: laboratory animals. *Rev. sci. tech. Off. int. Epiz.*, 24(2), 503-514.
- Bayer, K., Range, F., Aust, U., Steurer, M. & Huber, L. (2009). The touch-screen method as an implement for dog experiments. *Journal of Veterinary Behavior: Clinical Applications and Research*, 4(2), 51.
- Baymann, U., Langbein, J., Siebert, K., Nurnberg, G., Manteuffel, G. & Mohr, E. (2007). Cognitive enrichment in farm animals: the impact of social rank and social environment on learning behaviour in dwarf goats. *Berliner und Munchener Tierarztliche Wochenschrift*. 120(3-4), 89-97.
- Beerda, B., Schilder, M.B., van Hooff, J.A., de Vries, H.W. & Mol, J.A. (1999). Chronic stress in dogs subjected to social and spatial restriction. I. Behavioural responses. *Physiology and Behaviour*, 66(2), 233-242.
- Beirise, J., Reinhardt, V. (1992). Three inexpensive environmental enrichment options for group-housed *Macaca mulatta*. *Laboratory Primate Newsletter* 31 (1), 7–8.
- Bethell, E.J., 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, 185-195.
- Bloomsmith, M.A., & Lambeth, S.P. (1995). Effects of predictable versus unpredictable feeding schedules on chimpanzee behavior. *Applied Animal Behaviour Science*, 44, 65–74.
- Bloomsmith, M.A., Brent, L.Y. & Shapiro, S.J. (1991). Guidelines for developing and managing an environmental enrichment program for non-human primates. *Laboratory Animal Science*, 41, 372-377.
- Bloomsmith, M.A., Keeling, M.E. & Lambeth, S.P. (1990b). Videotapes: Environmental enrichment for singly housed chimpanzees. *Lab Animal*, 19, 42–46.

- Bloomsmith, M.A., Marr, M.J. & Maple, T.L. (2007). Addressing nonhuman primate behavioral problems through the application of operant conditioning: is the human treatment approach a useful model? *Applied Animal Behaviour Science*, 102, 205-222.
- Boissy, A., Arnould, C., Chaillou, E., Desire, L., Duvaux-Ponter, C., Greiveldinger, L., Leterrier, C., Richard, S., Roussel, S., Saint-Dizier, H., Meunier-Salaun, M.C., Valance, D. & Veissier, I. (2007a). Emotions and cognition: a new approach to animal welfare. *Animal Welfare*, 16, 37-43.
- Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, C., Dimitrov, I., Lanbein, J., Bakken, M., Veisser, I., & Aubert, A. (2007b). Assessment of positive emotions in animals to improve their welfare. *Physiology and Behaviour*, 92, 375-397.
- Branch, M.N., Nicholson, G. & Dworkin, S.I. (1977). Punishment-specific effects of pentobarbital: dependency on the type of punisher. *Journal of the Experimental Analysis of Behavior*, 28, 285-293.
- Briefer, E.F. & McElligott A.G. (2013). Rescued goats at a sanctuary display positive mood after former neglect. *Applied Animal Behaviour Science*, 146(1-4), 45-55.
- Brilot, B.O., Asher, L., & Bateson, M. (2010). Stereotyping starlings are more 'pessimistic'. *Animal Welfare*, 6, 32-36.
- Broom, D.M. & Johnson, K. (1993). *Stress and Animal Welfare*. Chapman & Hall: London, UK.
- Broom, D.M. (1983). Stereotypes as animal welfare indicators. In: D. Schmidt (Ed.) *Indicators Relevant to Farm Animal Welfare*, pp. 81-87. Martinus Nijhoff, The Hague.
- Brydges, N.M., Leach, M., Nicol, K., Wright, R. & Bateson, M. (2011). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, 81, 169-175.
- Burman O, Parker R, Paul E & Mendl M. (2008). A spatial judgement task to determine background emotional state in laboratory rats (*Rattus norvegicus*). *Animal Behaviour*, 76, 801-9.

- Burman, O., Parker, R., Paul, E., & Mendl, M. (2009). Anxiety-induced cognitive bias in non-human animals. *Physiology & Behavior*, *98*, 345–350.
- Carey, M.P. & Fry, J.P. (1995). Evaluation of animal welfare by the self-expression of an anxiety state. *Laboratory Animals*, *29*, 370-379.
- Carlstead, K. & Shepherdson, D. (1994). Effects of environmental enrichment on reproduction. *Zoo Biology*, *13*, 447-458.
- Celli, M.L., Tomonaga, M., Usono, T., Teramoto, M. & Nagano, K. (2003). Tool use task as environmental enrichment for captive chimpanzees. *Applied Animal Behaviour Science*, *81*, 171–182.
- Chen, H. & Wigand, R.T. (1999). Optimal experiences of web activities. *Computers in Human Behaviour*, *15*(5), 585-608.
- Clark, F.E. (2011). Great ape cognition and captive care: can cognitive challenges enhance well-being? *Applied Animal Behaviour Science*, *135*(1-2), 1-12.
- Coppinger, R. & Zuccotti, J. (1999). Kennel Enrichment: Exercise and Socialization of Dogs. *Journal of Applied Animal Welfare Science*, *2*(4), 281-296.
- Coppola, C., Grandin, T. & Enns, M. (2006). Human interaction and cortisol: Can human contact reduce stress for shelter dogs? *Physiology and Behaviour*. *87*(3), 537-541.
- Croney, C.C., Prince-Kelly, N. & Meller, C.L. (2007). A note on social dominance and learning ability in the domestic chicken (*Gallus gallus*). *Applied Animal Behaviour Science*, *105*(1-2), 254-258.
- Csatadi, K., Leus, K., & Pereboom, J. J. M. (2008). A brief note on the effects of novel enrichment on an unwanted behaviour of captive bonobos. *Applied Animal Behaviour Science*, *112*, 201-204.
- Csikszentmihalyi, M. (1990). *Flow: the psychology of optimal experience*. Harper Collins: New York.
- Dallaire, J.A., Meagher, R.K. & Mason, G.J. (2012). Individual differences in stereotypic behaviour predict individual differences in the nature and

- degree of enrichment use in caged American mink. *Applied Animal Behaviour Science*, 142(1-2), 98-108.
- David, H. A. (1987). Ranking from unbalanced paired-comparison data. *Biometrika*, 74, 432–436.
- Dawkins, M.S. (1983). Battery hens name their price: consumer demand theory and the measurement of ethological "needs". *Animal Behaviour*, 31, 1195–1205.
- Dawkins, M.S. (1988). Behavioural deprivation: A central problem in animal welfare. *Applied Animal Behaviour Science*, 20, 209-225.
- Dawkins, M.S. (1990). From an animal's point of view: Motivation, fitness and animal welfare. *Behavioral and Brain Sciences* 13, 1–61.
- De Vries, H., Stevens, J. M. G., & Vervaeke, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, 71, 585-592.
- Descartes, R. (1641) *Meditations on First Philosophy*. (as translated by Veitch, J.1901) accessed from: <http://www.wright.edu/cola/descartes/mede.html> on 2nd March 2010.
- Douglas, C., Bateson, M., Walsh, C., Bedue, A., & Edwards, S.A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Applied Animal Behaviour Science*, 139(1), 65-73.
- Doyle, R.E., Vidal, S., Hinch, G.N., Fisher, A.D., Boissy, A., & Lee, C. (2010). The effect of repeated testing on judgement biases in sheep. *Behavioural Processes*, 83(3), 349-352.
- Duncan, I.J.H. (1996). Animal welfare defined in terms of feelings. *Acta Agriculturae Scandinavica. Section A, Animal Science* 27, 29–35.
- Eilam, D., Zor, R., Szechtman, H. & Hermesh, H. (2006). Rituals, stereotypic behaviour and compulsive behaviour in animals and humans. *Neuroscience and Biobehavioral Reviews*, 30, 456–471
- Ellis, S.L. H. & Wells, D. L. (2008). The influence of visual stimulation on the behaviour of cats housed in a rescue shelter. *Applied Animal Behaviour Science*, 113, 166-174.

- Elmore, M.R.P., Garner, J.P., Kerr, A., Johnson, R.D., Kirkden, B.T.R., & Pajor, E.A. (2011). Getting around social status: Motivation and enrichment use of dominant and subordinate sows in a group setting. *Applied Animal Behaviour Science*, 133(3), 154-163.
- Enkel, T., Gholizadeh, D., von Bohlen und Halbach, O., Sanchis-Segura, C., Hurlemann, R., Spanagel, R., Gass, P. & Vollmayr, B. (2010). Ambiguous-cue interpretation is biased under stress- and depression-like states in rats. *Neuropsychopharmacology*, 35, 1008-1015.
- Eysenck, M.W., Mogg, K., May, J., Richards, A. & Mathews, A. (1991). Bias in interpretation of ambiguous sentences related to threat in anxiety. *Journal of Abnormal Psychology*, 100, 144-150.
- Fagot, J. & Paleressompoulle, D. (2009). Automatic testing of cognitive performance in baboons maintained in social groups. *Behaviour Research Methods*, 41(2), 396-404.
- Fox, C., Merali, Z., & Harrison, C. (2006). Therapeutic and protective effect of environmental enrichment against psychogenic and neurogenic stress. *Behavioural Brain Research*, 175, 1-8.
- Franks, B. (2012). *Engaging in effectiveness: highlighting the role of challenge in well-being and welfare* (PhD Thesis), Columbia University.
- Fraser, D. (2009). Animal behaviour, animal welfare and the scientific study of affect. *Applied Animal Behaviour Science*, 118(3-4), 108-117.
- Gammel, M.P., De Vries, H., Jennings, D.J., Carlin, C.M., & Hayden, T.J. (2003). David's score: a more appropriate dominance ranking method than Clutton-Brock *et al.*'s index. *Animal Behaviour*, 66, 601-605.
- Giraldeau, L.-A. & Lefebvre, L. (1986). Exchangeable producer and scrounger roles in a captive flock of feral pigeons: a case for the skill pool effect. *Animal Behaviour*, 34, 797-803.
- Giraldeau, L.A. & Lefebvre, L. (1987). Scrounging prevents cultural transmission of food-finding behaviour in pigeons. *Animal Behaviour*, 35, 387-394.

- Graham, L., Wells, D.L. & Hepper, P.G. (2005a). The influence of olfactory stimulation on the behaviour of dogs housed in a rescue shelter. *Applied Animal Behaviour Science*, 91, 143–153.
- Graham, L., Wells, D.L. & Hepper, P.G. (2005b). The influence of visual stimulation on the behaviour of dogs housed in a rescue shelter. *Animal Welfare*, 14(2), 143-148.
- Hagen, K. & Broom, D.M. (2004). Emotional reactions to learning in cattle. *Applied Animal Behaviour Science*, 85, 203-213.
- Hall, C.B., Lipton, R.B., Sliwinski, M., Katz, M.J., Derby, C.A. & Verghese, M.B. (2009). Cognitive activities delay onset of memory decline in persons who develop dementia. *Neurology*, 73, 316-361.
- Harding, E.J., Paul, E.S. & Mendl, M. (2004). Cognitive bias and affective state. *Nature*, 427, 312.
- Harley, H.E., Fellner, W. & Stamper, M.A. (2010). Cognitive research with dolphins (*Tursiops truncatus*) at Disney's The Seas: a program for enrichment, science, education and conservation. *International Journal of Comparative Psychology*, 23, 331-343.
- Hausberger, M., Gautier, E., Muller, C. & Jengo, P. (2007). Lower learning abilities in stereotypic horses. *Applied Animal Behaviour Science*, 107, 299-306.
- Haverbeke, C. Diederich, E. Depiereux & Giffroy, J.M. (2008). Cortisol and behavioural responses of working dogs to environmental challenges. *Physiology and Behavior*, 93(1–2), 59-67.
- Held, S.D.E. & Špinka, M. (2011). Animal play and animal welfare. *Animal Behaviour*, 81(5), 891-899.
- Hill, S.P. & Broom, D.M. (2009). Measuring zoo animal welfare: theory and practice. *Zoo Biology*, 28(6), 531-544.
- Home Office (2013) *Draft guidance on the Operation of the Animals (Scientific Procedures) Act 1986 (as amended)*. [Online] [Accessed on 12th March 2014] Available from: <http://www.gov.uk/government/uploads/system/uploads/attachment>

- Honess, P.E. & Marin, C.M. (2006). Enrichment and aggression in primates. *Neuroscience and Biobehavioral Reviews*, 30, 413–436.
- Hovland, A.L., Akre, A.K., Flø, A., Bakken, M., Koistinen, T. & Mason, G.J. (2011). Two's company? Solitary vixens' motivations for seeking social contact. *Applied Animal Behaviour Science*, 135(1–2), 110-120.
- Hubrecht, R. (1993). A comparison of social and laboratory environmental enrichment methods for laboratory housed dogs. *Applied Animal Behaviour Science*, 37, 345–361.
- Hubrecht, R. (1995). Dogs and Dog Housing. In: Smith, C.P. & V. Taylor (Ed.). *Environmental Enrichment Information Resources for Laboratory Animals: 1965 - 1995: Birds, Cats, Dogs, Farm Animals, Ferrets, Rabbits, and Rodents*. AWIC Resource Series No. 2. U.S. Department of Agriculture, Beltsville, MD and Universities Federation for Animal Welfare (UFAW), Potters Bar, Herts, UK, pp. 49-62.
- Hubrecht, R.C., Serpell, J.A. & Poole, T.B. (1992). Correlates of pen size and housing conditions on the behaviour of kennelled dogs. *Applied Animal Behaviour Science*, 34, 365–383.
- Ijichi, C.J., Collins, L.M. & Elwood, R.W. (2013). Evidence for the role of personality in stereotypy predisposition. *Animal Behaviour*, 85(6), 1145-1151.
- Inglis, I.R., Forkman, B. & Lazarus, J. (1997). Free food or earned food? A review and fuzzy model of contrafreeloading. *Animal Behaviour*, 53, 1171-1191.
- Jensen, M.B., Vestergaard, K.S. & Krohn, C.C. (1998). Play behaviour in dairy calves kept in pens: the effect of social contact and space allowance. *Applied Animal Behaviour Science*, 56(2-4), 97-108.
- Kalbe, C. & Puppe, B. (2010). Long-term cognitive enrichment affects opioid receptor expression in the amygdala of domestic pigs. *Genes Brain and Behaviour*, 9, 75–83.

- Kiley-Worthington, M. (1983). *The behaviour of beef suckler cattle (Bos taurus)*. Reihe Tierhaltung Bd. 14, Birkhäuser, Basel.
- Kirkden, R.D. & Pajor, E.A. (2006). Using preference, motivation and aversion tests to ask scientific questions about animals' feelings. *Applied Animal Behaviour Science*, 100, 29–47.
- Lang, P.J., Davis, M. & Ohman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of Affective Disorders*, 61, 137-159.
- Langbein, J., Nurnberg, G. & Manteuffel, G. (2004). Visual discrimination learning in goats and associated changes in heart rate variability. *Physiology and Behaviour*, 82, 601-609.
- Langbein, J., Siebert, K., & Nurnberg, G. (2009). On the use of an automated learning device by group-housed dwarf goats: Do goats seek cognitive challenges? *Applied Animal Behaviour Science*, 120 (3-4), 150-158.
- Latham, N. & Mason, G. (2010). Frustration and perseveration in stereotypic captive animals: is a taste of enrichment worse than none at all? *Behavioural Brain Research*, 211, 96–104.
- Laule, G. & Wittaker, M. (2001). The use of positive reinforcement techniques with chimpanzees for enhanced care and welfare. In L. Brent (Ed.), *Care and management of captive chimpanzees* (pp. 243-265). San Antonio, TX: American Society of Primatologists.
- Lea, S.E.G. (1978). The psychology and economics of demand. *Psychological Bulletin*, 85, 441–466.
- Lea, S.E.G. (1979). Alternatives to the use of painful stimuli in physiological psychology and the study of behaviour. *ATLA Abstracts*, 7(1), 20-21.
- Leavens, D.A., Aureli, F., Hopkins, W.D., & Hyatt, C.W. (2001). Effects of cognitive challenge on self-directed behaviors by chimpanzees (*Pan troglodytes*). *American Journal of Primatology*, 55(1), 1-14.
- Leighty, K.A. & Frigaszy, D.M. (2003). Primates in cyberspace: using interactive computer tasks to study perception and action in non-human animals. *Animal Cognition*, 6, 137-139.

- Leliveld, L.M.C., Langbein, J. & Puppe, B. (2013). The emergence of emotional lateralization: Evidence in non-human vertebrates and implications for farm animals. *Applied Animal Behaviour Science*, 145(1-2), 1-14.
- Leyland, C.M. & Honig, W.K. (1975). Maintenance of behaviour controlling the duration of discriminative stimuli. *Journal of the Experimental Analysis of Behaviour*, 24, 207-214.
- Lustig, C., Shah, P., Seidler, R. & Reuter-Lorenz, P.A. (2009). Aging, training and the brain: A review and future directions. *Neuropsychology Review*, 19(4), 504-522.
- Mandell, D.J. & Sackett, G.P. (2008). A computer touch screen system and training procedure for use with primate infants: results from pigtail monkeys (*Macaca nemestrina*). *Developmental Psychobiology*, 50, 160-170.
- Manser, C.E., Elliott, H., Morris, T.H. & Broom, D.M. (1996). The use of a novel operant test to determine the strength of preference for flooring in laboratory rats. *Laboratory Animals*, 30, 1-6
- Mantuffel, G., Langbein, J. & Puppe, B. (2009). From operant learning to cognitive enrichment in farm animal housing: bases and applicability. *Animal Welfare*, 18, 87-95.
- Markowitz, H., & Line, S. (1989). Primate research models and environmental enrichment. In: Segal, E. (Ed.), *Housing, Care, and Psychological Well-being for Laboratory Primates*, pp. 203-212. Noyes Publications, Park Ridge, NJ.
- Mason, G.J. & Latham, N.R. (2004). Can't stop, won't stop: is stereotypic behaviour a reliable welfare indicator? *Animal Welfare*, 13, 57-69.
- Mason, G.J. (1991). Stereotypies: a critical review. *Animal Behaviour*, 41, 1015-1037.
- Mason, G.J., Clubb, R., Latham, N. & Vickery, S. (2007). Why and how should we use environmental enrichment to tackle stereotypic behaviour? *Applied Animal Behaviour Science*, 102, 163-188.

- Mason, G.J., Cooper, J. & Clareborough, C. (2001). Frustrations of fur-farmed mink. *Nature*, 410, 35-36.
- 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.
- Meehan, C.L. & Mench, J.A. (2007). The challenge of challenge: Can problem behaviour solving opportunities enhance animal welfare? *Applied Animal Welfare Science*, 102, 246-261.
- Mendl, M. & Paul, E.S. (2004). Consciousness, emotion and animal welfare: insights from cognitive science. *Animal Welfare*, 13, 17-25.
- 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), 839-840.
- Mendl, M., Burman, O.H.P., Parker, R.M.A. & Paul, E.S. (2009). Cognitive bias as an indicator of animal emotion and welfare: emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science* 118, 161 - 181.
- Mertens, P.A. & Unshelm, J. (1996). Effects of group and individual housing on the behaviour of kennelled dogs in animal shelters. *Anthrozoos*, 9(1), 40-51.
- Milgram, N.W. (2003). Cognitive experience and its effect on age dependent cognitive decline in beagle dogs. *Neurochemical Research*, 28, 1677–1682.
- Moncek, F., Duncko, R., Johansson, B.B. & Jezova, D. (2004). Effect of environmental enrichment on stress related systems in rats. *Journal of Neuroendocrinology*, 16, 423-431.
- Myers, D.G. & Diener, E. (1995). Who is happy? *Psychological Science*, 6(1), 10-17.
- Nakao, T., Nakagawa, A., Nakatani, E., Nabeyama, M., Sanematsu, H., Yoshiura, T., Togao, O., Tomita, M., Masuda, Y., Yoshioka, K., Kuroki,

- T., & Kanba, S. (2009). Working memory dysfunction in obsessive-compulsive disorder: a neuropsychological and functional MRI study. *Journal of Psychiatric Research*, 43(8), 784-91.
- Neuringer, A.J. (1969). Animals Respond for Food in the Presence of Free Food. *Science*, 3903(166), pp. 399-401.
- Newberry, R.C. (1995). Environmental Enrichment: Increasing the biological relevance of captive environments. *Applied Animal Behaviour Science*, 44, 229-243.
- Ogura, T. (2012). Use of video system and its effects on abnormal behaviour in captive Japanese macaques (*Macaca fuscata*). *Applied Animal Behaviour Science*, 141(3), 173-183.
- Olsson, I.A.S. & Keeling, L.J. (2002). The push-door for measuring motivation in hens: laying hens are motivated to perch at night. *Animal Welfare*, 11, 11–19.
- Osborne, S. (1977). The free food (contrafreeloading) phenomenon: A review and analysis. *Animal Learning & Behavior*, 5(3), 221-235.
- Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., & Ballard, C.G. (2010). Putting brain training to the test. *Nature*, 465, 775-778.
- Palestrini, C., Minero, M., Cannas, S., Rossi, E. & Frank, D. (2010). Video analysis of dogs with separation-related behaviours. *Applied Animal Behaviour Science*, 124, 61-67.
- Palya, W.L. & Zacny, J.P. (1980). Stereotyped adjunctive pecking by caged pigeons. *Animal Learning & Behavior*, 8(2), 293-303.
- Panksepp, J. & Lahvis, G. (2011). Rodent Empathy and Affective Neuroscience. *Neuroscience and Biobehavioral Reviews*, 35(9), 1864-1875.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundation of Human and Animal Emotions*. New York: Oxford University Press, p. 480.
- Panksepp, J. (2011). Empathy and the laws of affect. *Science*, 334(6061), 1358-1359.

- Paul, E., Harding, E. & Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*, 29, 469-491.
- Pepperberg, I.M. (2004). Cognitive and communicative capacities of Grey parrots - implications for many species. *Animal Welfare*, 13, 203-208.
- Pietropaolo, S., Branchi, I., Chiarotti, F. & Alleva, E. (2004). Utilisation of a physically-enriched environment by laboratory mice: age and gender differences. *Applied Animal Behaviour Science*, 88(1-2), 149-162.
- Platt, D.M. & Nowak, M.A. (1997). Videostimulation as enrichment for captive rhesus monkeys (*Macaca mulatta*). *Applied Animal Behaviour Science*, 52, 139-155.
- Pomerantz, O., & Terkel, J. (2009). Effects of Positive Reinforcement Training Techniques on the Psychological Welfare of Zoo-Housed Chimpanzees (*Pan troglodytes*). *American Journal of Primatology*, 71 687-695.
- Pomerantz, O., Terkel, J., Suomi, S.J. & 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, 689–698.
- Poole, T. (1997). Happy animals make good science. *Lab Animal*, 31(2), 116-124.
- Puppe, B., Ernst, K., Schon, P.C., & Manteuffel, G. (2007). Cognitive enrichment affects behavioural reactivity in domestic pigs. *Applied Animal Behaviour Science*, 105, 75-86.
- Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorisation of natural stimuli by domestic dogs. *Animal Cognition*, 11(2), 339-347.
- Rault, J. (2012). Friends with benefits: Social support and its relevance for farm animal welfare. *Applied Animal Behaviour Science*, 136(1), 1-14.
- Redding, R.E., Morgan, G.A. & Harmon, R.J. (1988). Mastery motivation in infants and toddlers: Is it greatest when tasks are moderately challenging? *Infant Behavior and Development*, 11, 419-430.

- Rousseau, J-J. (1754). *Discourse on Inequality*, preface.
- Rumbaugh, D.M., Washburn, D., & Savage-Rumbaugh, E.S. (1989). On the care of captive chimpanzees: methods of enrichment. In: Segal, E. (Ed.), *Housing, Care and Psychological Wellbeing of Captive and Laboratory Primates* pp. 357–375. Noyes Publications, New Jersey.
- Rygula, R., Pluta, H. & Popik, P. (2012). Laughing Rats Are Optimistic. *PLoS One*, 7(12), e51959.
- Sale, A., Berardi, N. & Maffei, L. (2009). Enrich the environment to empower the brain. *Trends in Neuroscience*, 32(4), 233-239.
- Sanger, M.E., Doyle, R.E., Hinch, G.N. & Lee, C. (2011). Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. *Applied Animal Behaviour Science*, 131(3-4), 94-103.
- Schapiro, S.J. & Bloomsmith, M.A. (1995). Behavioral effects of enrichment on singly-housed, yearling rhesus monkeys: An analysis including three enrichment conditions and a control group. *American Journal of Primatology*, 35, 89–101.
- Schipper, L.L., Vinke, C.M. , Schilder, M.B.H. & Spruijt, B.M. (2008). The effect of feeding enrichment toys on the behaviour of kennelled dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, 114(1-2), 182-195.
- Sherwin, C.M. & Glen, E.F. (2003). Cage colour preferences and effects of home cage colour on anxiety in laboratory mice. *Animal Behaviour*, 66, 1085-1092.
- Sherwin, C.M. & Nicol, C.J. (1997). Behavioural demand functions of caged laboratory mice for additional space. *Animal Behaviour*, 53, 67–74.
- Sherwin, C.M. & Olsson, I.A.S. (2004). Housing conditions affect self-administration of anxiolytic by laboratory mice. *Animal Welfare*, 13, 33-38.
- Sherwin, C.M. (1996). Laboratory mice persist in gaining access to resources: a method of assessing the importance of environmental features. *Applied Animal Behaviour Science*, 48, 203–214

- Shields, S.J., Garner, J.P. & Mench, J.A. (2004). Dustbathing by broiler chickens: a comparison of preference for four different substrates. *Applied Animal Behaviour Science*, 87(1-2), 69-82.
- Shiverdecker, M.D., Schiml, P.A. & Hennessy, M.B. (2013). Human interaction moderates plasma cortisol and behavioral responses of dogs to shelter housing. *Physiology and Behavior*, 109, 75-79.
- Shuangying, Y., Zhigang, J., Hui, Z., Chunwang, L., Enquan, Z., Jinguo, Z., & Carin, H. (2009). Effects of odors on behaviors of captive Amur leopards *Panthera pardus orientalis*. *Current Zoology* 55(1), 20-27.
- Shyne, A. (2006). Meta-analytic review of the effects of enrichment on stereotypic behavior in zoo mammals. *Zoo Biology*, 25, 317–337.
- Smith, M.A., Iordanou, J.C., Cohen, M.B., Cole, K.T., Gergans, S.R., Lyle, M.A. & Schmidt, K.T. (2009). Effects of environmental enrichment on sensitivity to cocaine in female rats: importance of control rates of behaviour. *Behavioural Pharmacology*, 20(4), 312-321.
- Smith, M.E., Gopee, N.V, & Ferguson, S.A. (2009). Preferences of minipigs for environmental enrichment objects. *Journal of the American Association of Lab Animal Science*, 48(4), 391-394.
- Starling, M. (2012). A portable, automated apparatus for testing cognitive bias in dogs. *Proceedings of the AIAM Annual Conference on Animal Management*, p.40-46.
- Stephen, J.M. & Ledger, R.A. (2006). A longitudinal evaluation of urinary cortisol in kennelled dogs, *Canis familiaris*. *Physiology & Behavior*, 87, 911–916.
- Swaisgood, R.R. & Shepherdson, D.J. (2005). Scientific approaches to enrichment and stereotypies in zoo animals: what's been done and where should we go next? *Zoo Biology*, 24, 499-518.
- Tapkı, I., Şahin, A. & Onal, G. A. (2006). Effect of space allowance on behaviour of newborn milk-fed dairy calves. *Applied Animal Behaviour Science*, 99 (1-2), 12-20.

- Tarou, L.R. & Bashaw, M.J. (2007). Maximizing the effectiveness of environmental enrichment: Suggestions from the experimental analysis of behaviour. *Applied Animal Behaviour Science*, 102, 189-294
- Tarou, L.R., Kuhar, C.W., Adcock, D., Bloomsmith, M.A. & Maple, T.L. (2004). Computer assisted enrichment for zoo-housed orang-utans (*Pongo pygmaeus*). *Animal Welfare*, 13, 445-453.
- Taylor, K.D. & Mills, D.S. (2007). The effect of the kennel environment on canine welfare: a critical review of experimental studies. *Animal Welfare*, 16, 435-447.
- Titulaer, M., Blackwell, E.J., Mendl, M., Casey, R.A. (2013). Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs. *Applied Animal Behaviour Science*, 147(1), 149-158.
- Udell, M.A.R., Dorey, N.R. & Wynne, C.D.L. (2010). The performance of stray dogs (*Canis familiaris*) living in a shelter on human-guided object-choice tasks. *Animal Behaviour*, 79(3), 717-725.
- Van de Weerd, H.A., Van Loo, P.L.P., Van Zutphen, L.F.M., Koolhaas, J.M. and Baumans, V. (2006). Strength of preference for nesting material as environmental enrichment for laboratory mice. *Applied Animal Behaviour Science*, 55(3-4), 369-382.
- 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(1-2), 151-6.
- Vasconcellos, A. S., Adania, C.H. & Ades, C. (2012). Contrafreeloading in maned wolves: implications for their management and welfare. *Applied Animal Behaviour Science*, 140(1-2), 85-91.
- Vasconcellos, A.S., Guimaraes, M.A.B.V., Oliveira, C.A., Pizzutto, C.S. & Ades, C. (2009). Environmental enrichment for maned wolves (*Chrysocyon brachyurus*): group and individual effects. *Animal Welfare*, 18, 289-300.

- Vickery, S. & Mason, G. (2004). Stereotypic Behavior in Asiatic Black and Malayan Sun Bears. *Zoo Biology* 23, 409–430.
- Viggiano, E., Marshall-Pescini, S., Petak, I., Prato-Previde, E., & Natoli, E. (2009). Are cognitive skills good indicators of welfare in shelter dogs? *Journal of veterinary behaviour clinical applications and research*, 58, p.79.
- Walker, M.D., Duggan, G., Roulston, N., Van Slack, A. & Mason, G. (2012). Negative affective states and their effects on morbidity, mortality and longevity. *Animal Welfare*, 21(4), 497-509.
- Washburn, D. A., Hopkins, W. D., & Rumbaugh, D. M. (1989). Automation of learning set testing: The video-task paradigm. *Behavior Research Methods, Instruments, & Computers*, 21, 281-284.
- Washburn, D. A., Hopkins, W. D., & Rumbaugh, D. M. (1990). The effects of competition upon video-task performance in monkeys (*Macaca mulatta*). *Journal of Comparative Psychology*, 104, 115-121.
- Washburn, D. A., Hopkins, W. D., & Rumbaugh, D. M. (1991). Perceived control in rhesus monkeys (*Macaca mulatta*): Enhanced video-task performance. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 123-127.
- Watanabe, S. (2007). How animal psychology contributes to animal welfare. *Applied Animal Behaviour Science*, 106, 193-202.
- Wechsler, B. & Lea, S.E.G. (2007). Adaptation by learning: It's significance for farm animal husbandry. *Applied Animal Behaviour Science*, 108, 197-214.
- Wells, D. (2004). A review of environmental enrichment for kennelled dogs, *Canis Familiaris*. *Applied Animal Behaviour Science*, 85, 304-317.
- Wells, D.L. & Hepper, P.G. (2000). Prevalence of behaviour problems reported by owners of dogs purchased from an animal rescue shelter. *Applied Animal Behaviour Science*, 69, 55-65.
- Wells, D.L. (2009). Sensory stimulation as environmental enrichment for captive animals: a review. *Applied Animal Behaviour Science*, 118, 1-11.

- Wells, D.L., Coleman, D. & Challis, M.G. (2006). A note on the effect of auditory stimulation on the behaviour and welfare of zoo-housed gorillas. *Applied Animal Behaviour Science*, 100, 327-332.
- Wells, D.L., Graham, L. & Hepper, P.G. (2002a). The influence of length of time in a rescue shelter on the behaviour of kennelled dogs. *Animal Welfare*, 11(3), 317-325.
- Wells, D.L., Graham, L., Hepper, P.G. (2002b). The influence of auditory stimulation on the behaviour of dogs housed in a rescue shelter. *Animal Welfare* 11, 385–393.
- Wells, D.L., Hepper, P.G., Coleman, D. & Challis, M.G. (2007). A note on the effect of olfactory stimulation on the behaviour and welfare of zoo-housed gorillas. *Applied Animal Behaviour Science*, 106(1), 155-160.
- Wemelsfelder, F. (1997). Environmental challenge. In M.C. Appleby & B.O. Hughes (Eds.). *Animal Welfare*, pp. 35-48. UK: CABI.
- Wilkie, D.M., Mak, T. & Saksida, L.M. (1994). Pigeons' landmark use as revealed in a feature-positive discrimination, digitized landscape touchscreen paradigm. *Behavioral Processes* 32, 87-100.
- Wills, A.J., Lea, S.E.G., Leaver, L.A., Osthaus, B., Ryan, C.M.E., Suret, M.B., Bryant, C.M.L., Chapman, S.J.A. & Millar, L. (2009). A Comparative Analysis of the Categorization of Multidimensional Stimuli: I. Unidimensional Classification Does not Necessarily Imply Analytic Processing; Evidence From Pigeons (*Columba livia*), Squirrels (*Sciurus carolinensis*), and Humans (*Homo sapiens*). *Journal of Comparative Psychology*, 123, 4, 391-405. ISSN: 0735-7036
- World Organisation for Animal Health (2013) *Terrestrial Animal Health Code* Vol. I and II. [Online] [Accessed on 28th March 2014] Available from: http://web.oie.int/eng/normes/mcode/en_chapitre_1.7.1.pdf
- Yamanashi, Y. & Matsuzawa, T. (2010). Emotional consequences when chimpanzees (*Pan troglodytes*) face challenges: individual differences in self-directed behaviours during cognitive tasks. *Animal Welfare*, 19, 25-30.

- Yerkes R.M. & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.
- Yerkes, R.M. (1925) *Almost Human*, Century, New York.
- Young, R.J. (2003). *Environmental Enrichment for Captive Animals*. Blackwell, Oxford.
- Zaias, J., Queeney, T.J., Kelley, J.B., Zakharova, E.S. & Izenwarer, S. (2008). Social and physical environmental enrichment differentially affect growth and activity of preadolescent and adolescent male rats. *Journal of the American Association of Laboratory Animal Science*, 47(2), 30-34.
- Zebunke, M., Birger Puppe, B. & Langbein, J. (2013). Effects of cognitive enrichment on behavioural and physiological reactions of pigs. *Physiology and Behavior*, 118, 70–79.