Assessment of positive emotion in horses

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Veterinary Science

at Massey University, Manawatū, New Zealand

Rachael Bridget Stratton

2022

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Abstract

Objective, non-invasive indicators of the subjective experience of positive emotion are required to support assessment and improvement of animal welfare. Emotion is unique to the individual and indicators of emotion are indirect. The aim of this thesis was to ascertain if body and facial behaviours and physiological parameters reflected the emotional experiences of horses. Following review of the theoretical and experimental literature, three experiments were conducted, and an alternative emotional arousal-valence framework was proposed.

Based on the preferences of individual horses, the relative arousal level and emotional valence induced by four stimuli (wither grooming, motionless person, intermittent spray, and being left alone) were ordered. Behavioural and physiological parameters were then measured during exposure to each stimuli.

The indicators of contrasting affective experiences in horses were found to be heart rate, heart rate variability, eye temperature, and behaviours involving legs, neck, tail, ears, eyes, eyebrows, mouth, chin, and nares. Several behaviours differed across all three arousal levels or valence levels. Positive emotional valence was indicated by a decreased rate of neck very low, left ear forward, left or right ear back, blink, angled eyebrow, nares flared, nares neutral, and/or an increased rate of chin wobble, small eye aperture, or oral investigation behaviours. Higher arousal was indicated by an increased rate of neck very high, tail swishing, or higher odds of contracted lips, and/or a decreased rate of right ear forward or to the side behaviours. Reduction from 16 to six parameters may be possible. The findings may be used to aid interpretation of horse emotional experience and in the assessment and improvement of horse welfare.

The research approach and framework described in this research may be suitable for future research in horses and other species.

Acknowledgements

I sincerely thank the individuals and organisations that contributed to my studies and the development of this thesis.

I would like to recognise the fundamental catalysts for this work, the animals that enrich our lives, and our regard for their welfare. I would also like to acknowledge Churchill, the other horse subjects, and their caretakers, without which this research would not have been possible. Approval for the research was obtained from the Massey University Animal Ethics Committee for the experiments described in this thesis.

Thank you to my supervisors, Kevin Stafford, Ngaio Beausoleil, Naomi Cogger, Mairi Stewart, and Natalie Waran, for your guidance on this marathon. Your extensive and complementary knowledge, experience, perspectives, and humour have been formative and supportive in all aspects. I am also very grateful to Sarah Rosanowski for her indispensable knowledge, assistance, and support.

I am very thankful to the staff and students at Massey University and AgResearch who provided practical assistance throughout my PhD studies. To Rachel Munn, thank you for your positivity and good humour during assistance with 'Happy Horses'. Thank you to Jenny McCardle for her expertise in proofreading and formatting.

Furthermore, I gratefully acknowledge the financial support for this research from the New Zealand Government's Ministry for Primary Industries, the Building Research Capabilities in Strategically Relevant Areas (BRCSRA), and McGeorge Research funds.

A heartfelt thank you to all my friends, my parents, and my extended family that have listened with interest, facilitated my study, challenged, and encouraged me. Your contributions were vital to bringing this project to its conclusion. To Hayley Squance, my travelling companion on this metaphorical journey, thank you so much for believing in me.

To my best friend and husband, Scott, I offer my deepest gratitude for your unwavering love and support; you were my champion that bolstered and spurred me on.

Finally, I would like to dedicate this thesis to my daughters Ellandria and Rosielea. We are fortunate to come from a line of strong-minded, compassionate women. One day, I hope that you will understand why this was so important for me to accomplish, but now that it is done, yes, I absolutely can come and push you on the swings!

Perseverance [pur-s*uh*-**veer**-*uh*ns]

noun - continued effort to do or achieve something, even when this is difficult or takes a long time

Table of contents

List of figures

List of Appendix figures

List of tables

List of Appendix tables

List of abbreviations

Abbreviation	Term
ACTH	Adrenocorticotropic hormone
ANS	Autonomic nervous system
AV	Arterio-ventricular
bpm	Beats per minute
CNS	Central nervous system
CT	Computed tomography
ECG	Electrocardiography
EEG	Electroencephalography
ET	Eye-area temperature
FFT	Fast Fourier transformation
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy, measures oxyhaemoglobin concentration as an indicator of changes in cerebral perfusion and thus brain activity
HF	High frequency peak of heart rate variability spectral analysis
HPA	Hypothalamic-pituitary-adrenocortical axis
HR	Heart rate
HRV	Heart rate variability
IBI	Inter-beat interval, also known as NN or R-R
IRT	Infra-red thermography
LF	Low frequency peak of heart rate variability spectral analysis
LF/HF	Ratio of LF to HF. Also known as SNSI.
Min	Minute(s)
ms	Millisecond
PAW	Positive animal welfare
PET	Positron emission tomography
PNS	Parasympathetic nervous system
PQRST	Apices of the wave trace of a heartbeat on an electrocardiogram
PSA	Power spectral analysis

Abbreviation	Term
PSD	Power spectral density
RMSSD	Root mean square of successive differences of all IBI
RMSSD/SDNN	Ratio of RMSSD to SDNN
RR	Respiratory rate
R-R	Time interval in milliseconds between successive R waves of heart beats. Also known as IBI
SAM	Sympathetic adrenal medullary pathway
SDNN	Standard deviation of normal-to-normal IBI. Also known as SDRR.
sec	Seconds
SH	Humidity above the skin
SNS	Sympathetic nervous system
ST	Skin temperature
VLF	Very low frequency

Chapter 1 General introduction

1.1 The need for indicators of emotion in animals

Animal welfare is fundamentally about what animals experience and how coping with their environment impacts on them (Appleby et al. 2018). Emotion is one element of that experience. The ability to scientifically assess the emotional experience of animals is important for evaluation of animal welfare.

Historically, welfare standards have focused on the minimisation of negative emotional experiences, such as pain, fear, and distress (Yeates et al. 2008; Mellor 2016; Webb et al. 2018). Accordingly, most of the research on emotion in animals has centred on negative emotions (Boissy et al. 2007b). The ability to identify situations that elicit negative emotion allows animal caretakers to prevent or minimise them and thus reduce poor welfare (Kelly et al. 2021).

Animal welfare scientists now regard positive or optimal welfare, rather than just the absence of negative welfare, as a desired outcome (Boissy et al. 2007b; Yeates et al. 2008; Mellor et al. 2015; Lawrence et al. 2019). It can be argued that if an animal is experiencing frequent positive emotion, or is in a positive emotional state, then its needs are being met, and its welfare is good (Boissy et al. 2007b; Webb et al. 2018; Lawrence et al. 2019). There is little risk to welfare from encouraging opportunities for experience of positive emotion. However, if the evidence for positive emotion was disregarded and animals were assumed to be devoid of the capacity for pleasure, that may impact animal caretakers' interactions and management practices, presenting a greater risk to the quality as well as quantity of an animal's life (Yeates et al. 2008).

There is a lack of agreement on how to assess positive emotional experiences in animals. In recent decades, researchers have begun to explore the existence, detection, and measurement of positive emotions, such as pleasure. This is a developing but incomplete field (Boissy et al. 2007b; Keeling 2019; Kremer et al. 2020). Valid, objective, reliable, and preferably non-invasive indicators of positive emotion are needed to augment animal welfare assessment, promote positive welfare by providing opportunities for experience of positive emotion, and to improve human safety during handling and training (Hall et al. 2018).

1.2 Research objective

The aim of the research in this thesis was to ascertain if the emotional experiences of horses are reflected in their behaviour and physiology. Furthermore, the aim was to determine how particular behavioural and physiological parameters may indicate emotional valence and/or arousal in horses.

1.3 Species of interest

The horse, *Equus ferus caballus*, was chosen as the subject species for this body of work as horses are utilised for work, meat, sport/leisure, entertainment, and as pets (McGreevy 2012). As such, they may be both a companion animal and a production animal species. Anecdotally, people think they can tell when horses are happy and what horses like (invited audience feedback at the 6th International Society of Equitation Science Conference, Uppsala, Sweden, personal communication, July 31, 2010). However, horses, as a prey species, may be less likely to visually display emotion than a predator species in case it draws predatory attention (McGreevy 2012). A survey of equestrians concluded that although 94% of respondents claimed that they could tell when their horses were happy (described as “forward” or energetic when being ridden, or unresponsive to loud noises/scary objects), they were overconfident and incorrect (Bornmann et al. 2021). Thus, horses may be a more universal, but also challenging, domesticated species in which to study indicators of positive emotion.

It is suggested that if indicators of positive emotion can be found in horses, the research approach may be a useful method that could be adapted and applied to other species.

1.4 Structure of the thesis

To accomplish the research objectives, the knowledge and research on emotion in horses and other animals were reviewed. The scientific literature falls broadly into two areas, theoretical and experimental. The review presented in Chapter 2 discusses literature regarding theories on the existence and nature of emotions in animals and how emotion relates to animal welfare. Chapter 3 presents the experimental research on emotion in horses and other animals. Potential indicators of emotion and stimuli that may be used to induce positive or negative emotional responses in horses are discussed.

From this knowledge base, several experiments were conducted to investigate potential non-invasive indicators of emotion in horses. Chapter 4 describes the initial experiment that was conducted to measure behavioural and physiological responses to putatively positive and negative experiences. Based on the findings from this study, a new methodological approach requiring two further experiments was pursued.

In Chapter 5, a method of preference testing to identify an individual horse's preferred stimulus is presented. The final experiment used the findings from this study to investigate selected physiological, behavioural, and facial movement parameters as indicators of emotion in horses. Chapter 6 presents the general materials and methods of the final experiment. The more specific methods, results, and discussion relating directly to each parameter category are presented in the following three chapters: Chapter 7 presents physiological measures, Chapter 8 discusses gross body behaviours, and Chapter 9 covers head and facial movements. Chapter 10 brings together the results from Chapters 7-9, provides deeper interpretation of the results, and discusses and critiques the findings of the experiment. Multivariate analysis of the results of the experiment are presented in Chapter 11.

Finally, in Chapter 12, the contribution of the body of work as a whole is considered, conclusions are made, and further research directions are suggested.

Chapter 2 Review of literature on theories of emotion in animals

2.1 Introduction

This chapter provides background information and attempts to show that the foundational premise, that animals can experience positive and negative emotion, can be accepted. Following the introduction, key concepts in the study of emotion are defined. The scientific literature regarding theories on the existence and nature of emotion in animals is discussed with reference to anatomy and physiology, evolution, and philosophy. The purported functions of emotion are given, the types of emotions that may exist in animals are mentioned, and a framework for the study of emotion in animals is briefly discussed. Finally, the relevance of positive emotion to animal welfare is outlined.

The literature on emotions and emotional states in animals can be broadly divided into those that deal primarily with theoretical concepts and those that provide results from experimental research. While the scientific debate on the existence of emotion in animals, and the specific types of emotions they may be capable of feeling, dates back centuries, there is still no consensus on these topics (Fraser 2008).

2.2 Key concepts

2.2.1 Emotion

Emotion is an elusive concept to define (Paul et al. 2018). An emotion can be broadly defined as an innate, intense but short-lived response to an event that has behavioural, physiological, subjective (sometimes referred to as the feeling), and cognitive components (Panksepp 2005; Paul et al. 2005; Boissy et al. 2007b; Garland et al. 2010; de Waal 2011; Kremer et al. 2020). The components of emotion do not operate independently of each other, with bidirectional interaction or feedback possible in an iterative, dynamic process (Kremer et al. 2020). Whilst not universal, this componential approach appears commonly in the literature. Two factors stand out: the immediacy and the brevity of emotions.

The existence of the subjective and cognitive components of emotion in animals is subject to debate (Paul et al. 2005). For example, the view posed by Paul et al. (2005) categorised emotion in animals as being object (or stimulus) associated, that is, they do not result from internal stimuli. Animal sentience is closely related to the topic of emotion. Definitions of animal sentience vary in complexity from the simple ability to perceive sensation, to the capacity to experience emotion, or more deeply, to be able to be consciously aware of one's own emotions, and are also subject to debate.

2.2.2 Affect

In the context of emotion, the term affect can be taken to mean many things, such as approach and avoidance behaviour, which is observable in many simple organisms including plants, or as a term to encompass emotions and emotional states (Kremer et al. 2020; Paul et al. 2020). Sometimes the term is used to refer to the subjective component of emotion, or to a particular trait in an individual, such as personality. It can also refer to sensory feelings (e.g. heat, nociception) or homeostatic feelings (e.g. hunger, thirst) (Panksepp et al. 2017).

This can become confusing, especially when combined with common usage of the word affect, for example, 'fear is an emotion (affect) that has an influence (affect) on future responses to similar stimuli'. In this review, affect has the same meaning as emotion, and affective state has the same meaning as emotional state, as is commonly used in the literature. However, the use of affect, and the term feelings, will be minimised for clarity.

2.2.3 Behavioural component

Emotions can be characterised by the behaviours that accompany them. Behaviour may be defined as "the internally coordinated responses (actions or inactions) of whole living organisms (individuals or groups) to internal and/or external stimuli, excluding responses more easily understood as developmental changes" (Levitis et al. 2009). In the context of this thesis, behaviour is taken to be outwardly observable or audible changes in an animal. Behaviours may involve vocalisations, individual body or facial movements, and motor action patterns.

2.2.4 Physiological component

Emotions are accompanied by changes in physiological arousal, which is mainly mediated via the autonomic nervous system (Weiten 1992). There may be changes in heart rate, blood pressure, respiratory rate, piloerection, pupil size, sweat, catecholamine levels, and immunological and other neuroendocrinological changes. Many physiological variables have been investigated in studies of pain, fear, and distress in animals (Fraser 2008).

As well as general arousal, and the associated readiness for action, it has also been suggested that there may be physiological changes specific to different emotions. Thus, emotions could be characterised not only in terms of physiological arousal (low-high), but also emotion-specific physiological changes (Kremer et al. 2020).

2.2.5 Subjective component

The subjective component is also known as the conscious or experiential component, or the ‘feeling’ (Paul et al. 2020). In humans, emotion is subjective, personalised, may be valenced (experienced as positive/pleasurable or unpleasant/negative), and is unique to the being feeling it. Whilst emotion may be categorised, typed, and graded based on indicators such as self-report, physiological changes, behaviour, and cognitive appraisal, the exact way an emotion is perceived by a subject is intangible and irreproducible between individuals. This accounts for differences in the intensity and type of emotional response by individuals to the same stimuli, or in the same individual at different times. It may reflect the differences in experience, learning and memory, genetics, neural processing and neural plasticity, self-awareness, language and other communication skills, health status, and temperament that make subjects individual. Traits, such as temperament or personality, may influence emotional reactivity, or emotionality as it is sometimes referred to (Kremer et al. 2020). The existence or nature of the subjective component in animals is highly controversial, and it may vary with the relative cognitive complexity of a species (see Paul et al. 2005; Kremer et al. 2020; Paul et al. 2020). However, much empirical research is based on the assumption that the subjective component can be inferred from the other components of emotion (Mendl et al. 2010).

2.2.6 Cognitive component

The cognitive, or thought, component of emotion may involve a combination of perceptive, recall, evaluative, learning, and labelling processes. Evaluation (appraisal) results in broad classifications of the situation as pleasant or unpleasant, and a determination of the fitness cost/benefit (Weiten 1992; Kremer et al. 2020). According to appraisal theory (also known as the cognitive model), emotions result from an individual's evaluation of a situation, which may be conscious or sub-conscious, and involves cortical and subcortical structures. However, some researchers propose that the basic emotions occur without significant neocortical involvement (Panksepp 2005). It may be that the more complex the emotion is, the larger the cognitive or appraisal component, with more involvement of the neocortex.

2.2.7 Emotional state

Some authors have used the term emotional, or affective, state interchangeably with emotion. For clarity, the term emotional state is used in this thesis to refer to the longer, less intense, over-arching mood or outlook, for example, depression. Momentary emotion is event-related, whereas emotional state or mood is not linked to a particular stimulus or event. It is said to be 'free-floating', arising from the cognitive integration of experienced emotions (Panksepp 2010; Kremer et al. 2020). Emotional state may be regarded as a summary or moving average of momentary emotions, although there is a bidirectional relationship between the two (Kremer et al. 2020). It is suggested that research on positive emotion in animals is necessary to inform research on positive emotional states, which may include affective happiness (Boissy et al. 2007b; Webb et al. 2018).

Affective happiness

Within the concept of life satisfaction, also referred to as quality of life and happiness, a third category of emotional experience is suggested. Termed affective happiness, it is similar to mood in that it is not stimulus related. It represents an underlying, more stable and persistent type of emotional state than mood, and may reflect a dynamic summary of lifelong emotional experiences (Webb et al. 2018). The exact duration or temporal relationship between emotion, emotional state/mood, and affective happiness is not clear. The concept of affective happiness might, at least partially, apply to animals. The other component of human life satisfaction, cognitive happiness, involves evaluation of life against standards or expectations of how life should be (Webb et al. 2018). It is unknown whether the cognitive happiness component applies to animals as this has not been explored.

Research suggests that emotional state in humans may be a reflection of the balance, or ratio, of positive to negative emotional experiences during the medium to long term, which can be intentionally manipulated to improve wellbeing (Garland et al. 2010; Webb et al. 2018; Kremer et al. 2020). It is a concept that may be the subject of future research and discussion in relation to animal welfare.

2.3 Existence of emotions in animals

The debate over the existence of emotion in animals can be considered from biological, evolutionary, and philosophical perspectives.

2.3.1 Anatomy and physiology

If physiological and behavioural reactions analogous to those of humans experiencing a specific emotion occur in animals in response to a similar stimulus, then it may be that the animal is also feeling that emotion (Panksepp 2005; de Waal 2011). Negative emotions, such as pain and fear, have been generally accepted as emotions occurring in animals based on this premise. However, positive emotion in animals may not be as straightforward to correlate with humans, as what humans find pleasant may not be what animals find pleasant.

Aside from the neocortex, the neuroanatomy, neural circuitry, and neurochemistry associated with emotion in humans has been found to be similar in animals, with the involvement of cortical and subcortical structures, such as the prefrontal cortex, insula, cingulate cortex, amygdala, thalamus, nucleus accumbens, ventral pallidum, and periaque-ductal grey (Weiten 1992; Panksepp 2005; Burgdorf et al. 2006; Wager et al. 2008; Panksepp et al. 2017; Paul et al. 2020). Electrical stimulation studies using fMRI and/or PET neuroimaging in humans and animals have demonstrated remarkable similarities with positive and negative emotion (Weiten 1992; Panksepp 2005; Wager et al. 2008).

2.3.2 Evolution

Emotion may have evolved as a mechanism for animals to avoid harmful situations and take advantage of useful opportunities (Rolls 2000; Paul et al. 2005; Boissy et al. 2007b). The existence of emotions in animals is supported by homology to humans and theoretical evolutionary function (Clark 2010; Waller et al. 2020). It is thought that emotion pre-dates the existence of higher cognitive anatomy and functions, like thought (Weiten 1992; Bruce et al. 1995).

Positive emotion may confer a survival advantage via the acquisition of resources and the maintenance of social bonds (Boissy et al. 2007b). Transfer of positive emotion within groups enhances group stability (Hall et al. 2018). Positive emotions may be a form of primary positive reinforcement of behaviour that does not otherwise give immediate benefits to the performer, but continued performance has long-term survival advantages, akin to the hedonic theory of motivation in the evolutionary context (Keeling 2019). For example, exploration may have low immediate potential for attainment of the sought resource, but animals are motivated to perform these behaviours, and they appear to be inherently rewarding (Boissy et al. 2007b). Social cohesion is important to many species and energy is expended in social behaviour e.g., mutual grooming, play, and herding. These behaviours have direct functions of coat or skin health, practice of behaviours such as fighting, and ensuring access to mates. Group living also enhances individual survival via dilution of individual risk of predation, increased risk surveillance, knowledge of food and water locations, and in some species, shared care of offspring. Thus, affiliative behaviours are important to survival, and motivation to perform them may be maintained through positive emotion.

The benefits of emotion must outweigh the costs (e.g. in time and energy) in order for it to be a successful adaptation (Fraser 2008). Situations inducing negative emotions may be of greater importance to short term survival (e.g., predation) than many situations inducing positive emotions (e.g., social bonding). We might then expect more immediate and greater physiological and behavioural responses with negative emotion, and thus a greater cost, but also a greater benefit.

It follows that negative emotional responses may be easier to identify and differentiate than more subtle responses with positive emotions. For example, in horses, heart rate was found to increase by 19% on average with a fear inducing auditory stimulus (Christensen et al. 2005) but decrease by only 4.3% with wither grooming, a putatively pleasant experience (McBride et al. 2004).

2.3.3 Philosophy

Central to philosophical debate on the existence of emotion in animals is the subject of consciousness (Panksepp 2005; Paul et al. 2020). A mainstay of the argument against emotion in animals has been the view that emotion only exists where an individual is cognitively aware that they are experiencing an emotion (Boissy et al. 2007b).

There is some evidence, based on neurobiology and human subliminal exposure, that supports the occurrence of emotion without conscious involvement (Panksepp 2005; Kremer et al. 2020). However, in animals, the degree of self-awareness or consciousness, that is, the ability to distinguish internal stimuli from external stimuli, remains controversial (Broom 2010). Scientists have traditionally erred on the side of caution rather than overstating the mental capacities of animals. Although, more recent evidence indicates that animal consciousness is more involved than first thought, and it is very likely that some animals experience at least simple conscious thoughts and feelings, whether or not they are self-aware (Kirkwood et al. 2001; Griffin et al. 2004; Mendl et al. 2004; Beshkar 2008; Smith 2009; Broom 2010).

The cognitive capacity and level of consciousness in animals, and its necessity for the existence of emotion, or at least the subjective component of emotion, will likely continue to be debated (Paul et al. 2020). Boissy et al. (2007b) stated that affective consciousness probably exists in some animals, depending on cognitive capacity. However, Paul et al. (2005) were reluctant to make a decisive conclusion on consciousness and cognitive capacity, and stated that further research on cognitive appraisal and emotion is needed. A provisional working hypothesis, as suggested by Burgdorf and Panksepp (2006), is that mammals have basic forms of affective consciousness which are similar to humans, but that the existence of cognitive self-awareness of emotion (ability to think about our own emotions) in animals needs further investigation.

2.4 Functions of emotion

Emotion may be considered in terms of purported survival functions, such as resolving competing motivations, learning and memory, and social interaction.

2.4.1 Resolving competing motivations

The emotion system allows rapid combination and appraisal of various inputs (external and internal sensation) and competing motivations to act as a switch for behavioural and physiological responses. Emotion may have arisen to allow an adaptive, appropriate, and efficient response to situations where there is potential threat (avoidance) or potential opportunity (approach), and a fitness cost of inaction, without the need for conscious thought and associated time delay (Paul et al. 2005; Boissy et al. 2007b; Garland et al. 2010).

2.4.2 Learning and memory

Emotion underpins learning and memory (Rolls 2000). Something may be rewarding because it induces positive emotion (positive reinforcement) or decreases negative emotion (negative reinforcement) (Panksepp 2010). Thus, behaviour that is beneficial to the animal is encouraged. The converse, with punishment, also applies. Emotion may have evolved alongside the cognitive tasks of learning and memory, with punishment avoidance and reward seeking behaviour seen in most animal species (Paul et al. 2005). Many animals, including the horse, display rapid and long lasting memory for situations that are potentially harmful or fear evoking, thus protecting the individual in the future (McGreevy 2004; Hanggi et al. 2009b).

2.4.3 Social interactions

Through communication to others, emotions may function to enhance individual, kin, and/or group survival. Rearing of offspring, social safety, hunting success, hierarchy, and bonding are assisted by communication of emotion (Fraser 2008). In social species, emotion can spread between animals in a process similar to emotional contagion (Kremer et al. 2020). For example, in a dangerous situation, it may be more important to quickly convey the emotional response of fear rather than elaborate on the cause of the fear. However, it may also be advantageous to mask communication of emotion, especially in prey species, when emotion may be associated with vulnerability to predation or social outcast (Hall et al. 2018).

The damaging effects of chronic stress may be alleviated by positive emotions (Garland et al. 2010). In many social species, the presence of peers, especially those that are less negatively affected, appears to buffer negative emotion (termed social support or social buffering) (Guesgen et al. 2014; Kremer et al. 2020).

2.5 Types of emotion

There is much debate over the number and types of emotions (and emotional states) that animals, or even humans, experience. As well as differing in duration (momentary emotion/emotional state) and intensity (strength), emotions may be considered as occurring discretely, as in a specific emotion e.g. fear, or as continua along dichotomous axes e.g. happy-sad (Weiten 1992; Boissy et al. 2007b; Kremer et al. 2020). Sandem, Braastad and Bøe (2002) suggest the existence of a frustrated-contented emotional axis in cows.

Emotion may be thought of by level of complexity. There is consensus for five basic, discrete emotions in humans: anger, fear, disgust, sadness, and joy. It is generally agreed that there are two levels of human emotions, with secondary emotions resulting from the interaction of primary emotions (Weiten 1992; Clark 2010). However, it has been theorised that three tiers of emotion exist that increase in complexity from homeostatic drivers to socially influenced emotion (Damasio et al. 2000; Kremer et al. 2020). Slightly differently, Panksepp (2005) describes primary, secondary, and tertiary process emotional systems. Whether, or to what extent, the levels, in either schema, involving more complex neo-cortical cognitive involvement exist in each animal species is debateable (Panksepp 2005).

How specific emotions are defined and whether they are experienced by animals is contentious. Examples of the types of discrete emotions that have been suggested in animals are pain, fear, anxiety, distress, pleasure, joy, sadness, rage, lust, boredom, pride, guilt, grief, comfort, curiosity, frustration, contentment, and empathy. Panksepp (2005) provides some neuroanatomical support from deep brain stimulation studies for mammals having seven primary, or core, emotional or motivational realms: seeking, fear, rage, lust, care, panic, and play. Stimulation of specific areas of the brain is said to result in expression of these emotions in similar ways to humans. Though, that is controversial (Panksepp et al. 2017).

The most experimental work has been conducted on pain and fear/anxiety. Some work has been done on more complex emotions, for example frustration (e.g. Sandem et al. 2002; Greiveldinger et al. 2011), and empathy (Panksepp et al. 2011; Guesgen 2015). Less has been done on positive emotions, for example pleasure, happiness, and positive anticipation.

Pain has been the subject of much debate and empirical research, with attitudes towards pain in animals having changed remarkably in the last century (Fraser 2008). Pain used to be regarded as a sensation only and not an emotion in animals (Bermond 2001). Pain is defined as "*an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage*" (Loeser 2011). Unlike other emotions, such as fear or happiness, pain can involve specific peripheral receptors. At a basic level, nociception, the detection of noxious stimuli by specific receptors, can be demonstrated by spinal reflex arcs, evidenced by the withdrawal reflex, in the absence of any cortical input. However, it has been shown in humans and animals that pain, actual or anticipated, stimulates neural pathways in the brain associated with the limbic system, as do other emotions (Hamra et al. 1993; van Oostrom et al. 2007).

Aside from pain and fear, given the lack of agreement on the specific types and definitions of emotions in humans, classification of other specific types of emotions experienced by animals seems premature. Fundamental agreement on the definitions of specific emotions and the methods by which to identify and measure them may be a prerequisite. At this point, consideration of emotion in broad categories of positive (pleasant, desirable) and negative (unpleasant, undesirable) valence may allow progress of experimental research on emotion in animals without controversial labelling of specific discrete emotions.

2.6 Frameworks for studying emotions in animals

The theories of tiered emotional complexity, and Panksepp's purported 7-basic-emotion-circuits model, still require animal researchers to interpret and classify their data into discrete emotions, or discrete categories of emotion (Panksepp et al. 2017). The lack of agreement on specific types of discrete emotions (and emotional states) and gold standards for evaluation makes direct comparison of studies purporting to measure specific emotions very difficult, as is also the case in study of human emotion (Mauss et al. 2009). Frameworks for broader consideration of emotion may help to resolve this issue and contextualise individual pieces of research.

Emotion can be described in terms of physiological arousal (high/low activation) and valence (positive/pleasant or negative/unpleasant) (Paul et al. 2005). The arousal versus valence emotional framework, also known as dimensional or core affect space, proposed by Mendl et al. (2010) (Figure 1), is very similar to an earlier adaptation of previous work described by Knutson et al. (2002) (Figure 2) and has its basis in the circumplex model of affect studied by Russell (1980). Arising from comparison of three different techniques for scaling emotion labels and principal component analysis of subject self-report, the model was originally proposed to reflect both the actual structure of human emotional experience, and the cognitive conceptualisation of emotion by humans (self-awareness, as required for self-report).

Mendl et al.'s (2010) framework provides a method of classification of discrete emotions into four quadrants. It allows for the two-dimensional interaction of arousal and valence along continua from relaxed to excited on the vertical axis, and negative to positive emotional valence on the horizontal axis. This may be a useful, although simplistic, framework to adopt whilst our knowledge of the specific types of emotions and their interrelationship in animals is developing, as it avoids the need for theoretical debate about types of discrete emotions. Instead of labelling an animal's emotional response as fear, it could be referred to as a quadrant four (or High Negative) emotion.

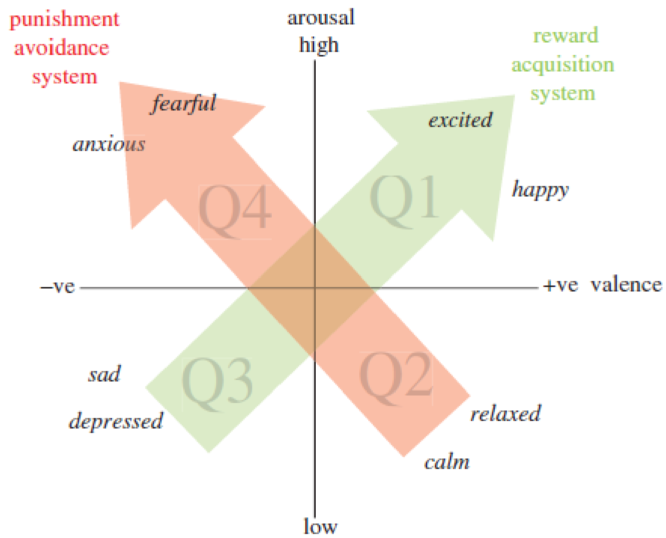


Figure 1 From Mendl et al. (2010). Core affect represented in two-dimensional space. Words in italics indicate possible locations of specific reported affective states (including discrete/basic emotions). Positive affective states are in quadrants Q1 and Q2, and negative states in quadrants Q3 and Q4. Arrows indicate putative biobehavioural systems associated with reward acquisition and the Q3–Q1 axis of core affect (green), and punishment avoidance and the Q2–Q4 axis of core affect (red). Adapted from Russell (e.g., Russell & Barrett 1999) and Panksepp (e.g., Burgdorf & Panksepp 2006).

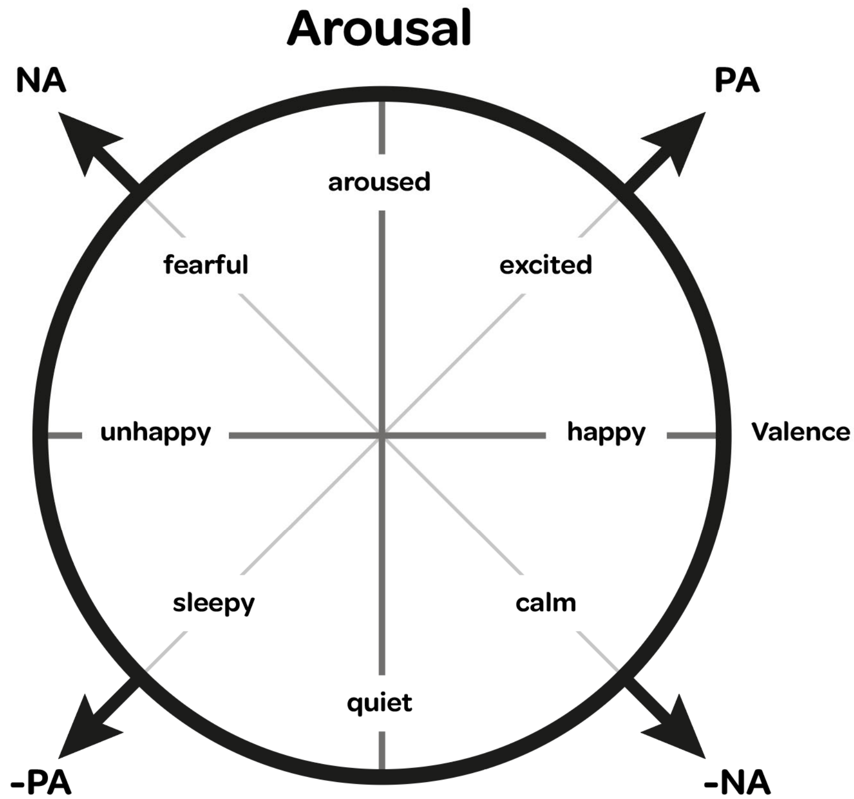


Figure 2 From Knutson et al. (2002). A two-dimensional map of affective space. Valence runs from negative (left) to positive (right), whereas arousal runs from high (top) to low (bottom). PA = positive activation; NA = negative activation. Adapted from "The Two General Activation Systems of Affect: Structural Findings, Evolutionary Considerations, and Psychobiological Evidence" by D. Watson, D. Wiese, J. Vaidya, and A. Tellegen, 1999, Journal of Personality and Social Psychology, 76.

2.7 Positive emotion and animal welfare

Emotion and welfare are inextricably linked. According to Fraser's (2008) recount on the history of animal welfare, society believes in the existence of emotion, both pleasant and unpleasant, in animals. Many people show concern for animal welfare because they care about how their animals, and animals in general, feel (Mendl et al. 2004; Keeling 2019).

Evolution of animal welfare science has continued over the last 20-30 years with a paradigm shift towards the concept of a welfare continuum from negative, through neutral (a theoretical point), to positive animal welfare (PAW) (Lawrence et al. 2019). Good welfare is now defined not only as an absence or minimisation (Mellor 2016) of negative welfare impacts, but where an animal is coping well with its environment *and* having positive experiences over the medium to long term (Keeling 2019). The current theme is provision of a good quality of life through the promotion of positive emotional experiences (Boissy et al. 2007a; Boissy et al. 2007b; Fraser 2008; Yeates et al. 2008; Balcombe 2009; Green et al. 2011; Mellor 2015; Webb et al. 2018; Keeling 2019; Lawrence et al. 2019). However, there is debate about the definition of the various terms involved, such as quality of life, good welfare, positive welfare, optimal welfare, 'life worth living', and the temporal associations (short-, medium-, long-term, whole of life).

The minimisation of negative states is understandably more compelling and has therefore been the focus of much of the past research on emotion and welfare indicators (Webb et al. 2018; Keeling 2019). Animal welfare science is now exploring the existence, detection, and measurement of positive emotions, such as pleasure, in order to inform discussion and provide methods for practical application (Boissy et al. 2007b; Reefmann et al. 2009a; Keeling 2019). It can be argued that if an animal is experiencing pleasure or happiness, or is in a happy emotional state, then its needs are being met, and its welfare is good (Boissy et al. 2007b).

The concept of positive welfare has become more apparent in welfare assessment models, such as in the European Welfare Quality scheme for quality assurance in the animal food chain, where the 12th criterion is positive emotional state (Blokhuis et al. 2010). However, this criterion has not made it into the practical application of assessment, presumably due to a lack of reliable, valid, and importantly, feasible indicators of it (Canali et al. 2009). Similarly, the updated Five Domains model for animal welfare assessment supports consideration of how adequacies or inadequacies in the four physical domains (nutrition, environment, health, and behaviour) may impact the emotional experience of an animal, in the fifth domain termed mental state (Mellor et al. 2015).

Recently, Webb et al. (2018) extended the concept of affective happiness, but not necessarily cognitive happiness, to animals, and suggested that the frequency and simple ratio of positive to negative emotions experienced (affective balance) is important in assessment of overall lifetime welfare. The validity of the affective balance concept is supported by Mellor et al. (2015), although they equate the term quality of life with animal welfare status, either at a single point in time, or over time. The idea of measuring the number and valence of emotional experiences, regardless of type, intensity, or duration, to calculate a ratio, above which life is deemed good, has an attractive simplicity that warrants further exploration.

2.8 Conclusion

Based on evolutionary benefit, analogous and homologous anatomy, physiology and behaviour, and philosophical arguments, it is plausible that animals experience emotion, including positive emotion. Whether or not, or to what degree, they are cognitively aware (conscious) is debateable. Other issues, such as the exact degree of consciousness and cognitive ability of each species compared to humans, whether emotion has a subjective component in animals, specific types of emotion experienced by animals, and which emotions involve conscious cognitive input and when, continue to be subjects of debate.

Adoption of the concept that minimisation of negative emotion and provision of positive emotion inducing experiences is necessary to protect an animal's welfare may allow for practical application of research on animal emotion. To progress experimental research on the detection and measurement of indicators of positive emotion in animals, several premises, for which there is some scientific support, must be accepted. These include that non-human mammals can experience emotion; that emotion experienced by animals can be positive or negative and can vary in level of arousal; and that variation in emotional valence is likely to result in behavioural, physiological, and cognitive changes, from which the subjective component may be inferred.

Chapter 3 Review of experimental literature on assessment of positive emotion in animals

3.1 Introduction

The aim of this review is to describe the parameters and emotion inducing experiences that have been investigated in previous studies of emotion in animals, and to identify methodological issues that may be applicable to the study of positive emotion in horses.

The identification and assessment of subjective experience in animals has been referred to as 'the hard problem' (Kremer et al. 2020). Although review of the literature on the existence of emotion in animals (Chapter 2) concluded that it is likely that they experience positive and negative emotion, it remains an assumption rather than a certainty. The subjective experience of emotion is unique to each individual; it cannot be quantified directly. Just as we cannot know the mind of another human being, we are unable to know the mind of an animal. Delving deeper into how animals experience emotion involves using indirect measures as proxies of the subjective experience.

Animals cannot verbalise the nature of their subjective experience of emotion, so self-report cannot be used as a gold-standard for identification and measurement of emotion (Boissy et al. 2007b). In humans, self-report is linked to observable changes in physiology (e.g., cardiac, neuroendocrine, brain activity, immune system), behaviour (e.g., body language, facial expression, vocalisation), and cognitive function (e.g., decision making). In animals, it may be that it is possible to infer emotional valence and/or arousal based on changes in physiology, behaviour, and, mainly for emotional states, cognition (Boissy et al. 2007b).

Objective and preferably non-invasive (or less-invasive) indicators of positive emotion in horses are desired in order to assess welfare status and welfare improvement efforts (e.g. environmental enrichment), and to promote positive welfare (Waran et al. 2017; Keeling 2019). Despite the lack of clear agreement on approaches to animal emotion, terminology, definitions, or indeed what specific emotions animals are capable of, much progress has been made on assessment of negative emotions such as pain and fear in many species. Research on assessing positive emotion is in its infancy but is gaining momentum (Boissy et al. 2007b; Webb et al. 2018; Kremer et al. 2020). Positive emotional stimuli may be less salient than negative ones (Smith et al. 2016), and accompanied by more subtle changes, thus identifying indicators of positive emotion may be more difficult (Whittaker et al. 2019).

This review focuses mainly on positive emotion, providing an overview of findings and a critique of methods. The literature on positive emotion in horses is limited, so the review includes literature on other species. Potential indicators of emotion fall into three categories: physiological, behavioural, and cognitive. Influential parameters are discussed in detail. This review also examines stimuli that may induce emotion in animals for research purposes.

3.2 Key concepts

3.2.1 Indicators of emotion

Emotion, or at least the subjective and cognitive components of it, cannot be directly measured, but can be inferred via the use of physiological and behavioural indicators. An indicator is a reliable variable that is generally agreed by scientists to be valid for making inferences by proxy about the parameter of interest (Keeling 2019). Since there is no gold standard to refer to, no indicator is perfect, and indicators may be subject to interpretation, it may be advisable to utilise several indicators to improve confidence in the inference. However, that requires careful interpretation, especially as indicators may not always appear to agree with each other.

It is important that indicators are valid (measure what they purport to), reliable (consistency of results), feasible (time, cost, practicality), minimally invasive (welfare, confounding), and animal based (observable, applicable) (Keeling 2019).

3.2.2 Emotion inducing experiences

To gather data on positive emotion in the research setting, it is necessary to provide the subjects with an experience likely to induce an emotion in them. Verification of the likely valence of the induced emotion is required, using means independent from the potential indicators being investigated. Otherwise, a circular argument may ensue.

Alternatively, studies could be conducted on spontaneously occurring episodes of putatively positive experiences such as play, mutual grooming, and exploration (approach behaviour) in animals whose welfare is not compromised and where the necessary resources are available (e.g., conspecifics and social play). However, data collection in this type of study is time consuming and may be limited to observational data. Without comparison to a negative emotional experience, only correlations but not causality can be investigated, and generalisation may be limited. In addition, control of conditions and confounds may be difficult in the natural setting.

3.3 Negative experiences and emotion in horses

Much of the literature on emotion in horses focuses on horse-human interaction or horse temperament (Christensen et al. 2005; Des Roches et al. 2008; Lansade et al. 2008c, b; Schmidt et al. 2010). Fear is one of the more commonly researched emotions in horses (Forkman et al. 2007).

The following markers have been used to measure negative emotions such as pain, stress and frustration: behaviour, heart rate (HR), respiratory rate (RR), rectal temperature, pupil diameter, heart rate variability (HRV), observer score (e.g. composite pain scale), cortisol (plasma, salivary, urinary, faecal), oxidative stress, adrenaline, and noradrenaline levels (Hamra et al. 1993; Taylor et al. 2002; Rietmann et al. 2004a; Rietmann et al. 2004b; Ashley et al. 2005; Dutton et al. 2009; Hellhammer et al. 2009; Lerche 2009; van Loon et al. 2010; Whitaker et al. 2011; Hausberger et al. 2012; Love 2012; Young et al. 2012; Lesimple 2020). Those that could be considered for their utility in measurement of positive emotional valence are discussed in section 3.5 - Assessment of positive emotion.

In negative emotion inducing experiences, sympathetic nervous system (SNS) tone increases and parasympathetic nervous system (PNS) tone decreases, however, cortisol may remained unchanged (Bachmann et al. 2003).

3.4 Potential emotion-inducing stimuli

There are a limited number of studies that have explicitly focused on positive emotion in horses. However, examination of studies on the behavioural and physiological responses of horses and other species undergoing putatively pleasant experiences may also help direct future work on markers of positive emotion in horses. For example, Feh et al. (1993), Normando et al. (2003), McBride et al. (2004) and Normando et al. (2007) considered the effects of mimicking allogrooming by humans on HR, but did not examine emotional context directly. Kato et al. (2003) used measures of HRV to suggest that warm water immersion was relaxing for horses, but did not mention emotion explicitly. Several others have examined the human-horse relationship (Hausberger et al. 2008; Fureix et al. 2009; Sankey et al. 2010b; Scopa et al. 2020). The treatments that were used to improve the response of horses to humans may have been rewarding and so induced positive emotion in the horses.

Inducement of positive emotion in animals in an experimental setting may be difficult. Individual variation in previous experience and temperament may impact whether an experience is positive or not. The experimental process, which often involves human manipulation, e.g. handling and restraint, may also confound either negatively (induce negative emotion) or positively (induce positive emotion), depending on how it is perceived by the individual and their level of habituation (Guesgen et al. 2017). It is suggested that, instead of inducing positive emotion in a research setting, it may be better to utilise naturally occurring events where the animal chooses to engage in the activity, such as interaction with objects/toys, grazing, and the use of automated grooming brushes.

Subjective emotional response is unique to the individual. What one finds pleasant, another may not. Care must be taken to independently determine that a putatively positive (or negative) stimulus is indeed regarded as pleasant (or unpleasant) by each subject. Without some form of self-report (as in humans), there is real risk of a tautology or circular argument ensuing. For example, where grooming by a human is used as a positive emotion inducing stimulus, and it is said to be positive because of the horse's behaviour (e.g., leaning into the grooming, head lowered), but then the horse's behaviour is also used as an outcome of the experiment, i.e., an indicator of positive emotion.

When selecting stimuli to use, many factors, such as the social and ethical responsibility not to cause unreasonable or unnecessary pain or distress, repeatability and standardisation, ease of use, the duration of the stimulus and response, likely arousal level, physical and mental health risks, and potential confounding of variables to be measured, need to be considered. Ideally, positive and negative emotion inducing experiences should be similar in terms of arousal or intensity, be based on the same substrate, and studies should include a 'neutral' or intermediate treatment for comparison.

3.4.1 Human-animal interaction

Human-animal interaction may be a suitable stimulus for inducing emotion in horses. Reviews of human-horse interactions discussed factors influencing how humans are perceived by horses and concluded that the emotional valence of interactions is important in horse welfare and human safety (Hausberger et al. 2008; Kelly et al. 2021). The response of horses to humans, termed a reactivity to humans temperament trait by Lansade et al. (2008a), was found to be generalisable between familiar and unfamiliar people, stable across time, and measurable from eight months of age. Human-horse interaction tests may assess behaviours of avoidance (e.g., fear, aggressive responses, withdrawal), indicative of negative experiences, and approach, investigative, or proximity behaviours (e.g. voluntary animal approach), indicative of pleasant experiences (Fureix et al. 2009; Sankey et al. 2010b; Lansade et al. 2019; Scopa et al. 2020; Kelly et al. 2021).

The valence of human-animal interactions may be altered through the processes of counterconditioning and desensitisation. It may involve exposure to a motionless human, tactile stimulation, for example stroking or brushing, and/or palatable food. In goats, quail, and horses, the behaviour of the dam towards humans impacts their offspring's behaviour towards humans. How a horse regards a human can be manipulated, using food and soft brushing to improve the relationship with the mare, and thus improve handling and training of foals (Henry et al. 2005). The effect was shown to be greater and longer lasting than the presence of a passive human or forced handling ("imprint" training, stroking, pushing to teat, haltering) of the foal.

Smith et al. (2018) found that horses preferred a human with a 'submissive' posture (slouching, hunched shoulders, feet together, relaxed knees, hands in front, 'closed' posture) to a 'dominant' posture (standing tall, feet hip-width apart, shoulders squared, chest out, hands by sides, 'open' posture). Horses that had been pre-conditioned to expect a food reward from approaching a human with a neutral posture were subject to a voluntary animal approach test with two human experimenters differing mainly in posture. In the first of the four test trials, more horses chose the submissive posture. The criterion for display of a preference was set at ≥ 75%, i.e., display of the same choice in at least 3 out of 4 test trials. Twenty-three of 29 horses preferred the submissive, six did not meet the preference criterion (no clear preference), and none preferred the dominant posture. No effect of horse age or sex on choice was found. Authors suggest that familiarity with the human experimenter was not necessarily required. However, horses had been conditioned to received rewards from the experimenters and no analysis of experimenter preference was able to be performed (due to the large number of demonstrators relative to subjects).

Evidence exists for emotional transfer from humans to animals using visual, acoustic, and olfactory stimuli in dogs and horses (Lanata et al. 2018). Smith et al. (2016) showed that horses differentiate between images of happy and angry male human faces (validated using the human Facial Action Coding System FACS). They concluded that the left gaze bias (right cerebral hemisphere) and more rapid increase in HR with the angry human facial expression demonstrates that the horses interpreted it as negatively valenced. There was no lateral bias demonstrated for the happy facial expression. It is unknown whether the findings are evolutionary based (innate behaviour, representing a conserved ability for recognition of emotional cues across species) or due to learning (prior association of an angry face with an unpleasant outcome). Studies on animal emotion that involve humans should consider the possible effects, intended or otherwise, of emotional transfer.

3.4.2 Tactile: stroking, scratching, massage, grooming

Allogrooming is suggested to be pleasurable and rewarding so that the behaviour is maintained over short and long terms. Perhaps in mimicry of allogrooming, tactile stimulation performed by a human, including stroking, massage, brushing/grooming, 'tickling', or scratching, has been used as a putatively positive experience in many species, including cattle, sheep, rats, primates, and horses (Feh et al. 1993; Normando et al. 2003; McBride et al. 2004; Burgdorf et al. 2006; Normando et al. 2007; Reefmann et al. 2009a; Yamamuro et al. 2010; Laister et al. 2011; Proctor et al. 2015; Serrapica et al. 2017; Janczarek et al. 2018b; LaFollette et al. 2018; Lansade et al. 2018; Tamioso et al. 2018; Lange et al. 2020). Cows will choose to engage in human stroking and attempt to initiate it. In cows, it has been associated with a reduction in cortisol, flight distance, fear of humans, and changes in HR (Proctor et al. 2015; Lange et al. 2020). Although allogrooming is not usually observed in sheep outside of maternal care, human tactile stimulation has been used to induce putatively positive emotion (Reefmann et al. 2009a). Tactile stimulation has been used successfully to habituate rats, via counterconditioning, to human handling and research procedures such as injections, indicating that it is perceived as rewarding (LaFollette et al. 2018).

Stroking by a human of generally preferred body areas was used to induce a putatively low arousal positive emotion in cows (Proctor et al. 2015). The stroking was performed for a total of 5 minutes per session on the withers, neck, forehead, and cheek areas by a familiar person wearing canvas gloves. The rate of 40-60 strokes per minute was the same as that observed when cows allogroom each other. The sessions were aborted if the cow moved away, stood up, or lay down during stroking, or if it started eating, or displayed aggressive or mounting behaviour during the 15-minute observation period.

In horses, grooming by humans of the withers/neck area may be pleasant or unpleasant (Lansade et al. 2019). Feh et al. (1993), following observations of mutual grooming in feral horses, reported a decrease in HR in response to grooming by a human. The authors suggested that the effect on HR is via parasympathetic stimulation, as the stellate ganglion, which has cardiac efferents, is located close to the preferred grooming site in horses. They concluded that grooming by a human at a preferred site (cranial wither area), with a movement frequency and duration similar to that observed during allogrooming (2/second for 3 minutes), is calming for the recipient, and may be rewarding. Although there was an effect of habituation, comparable but less striking results were found by Normando et al. (2003) on horses used for riding.

Similarly, McBride et al. (2004), in a study of massage-like grooming in riding school horses, found that grooming of the wither and mid-neck areas decreased HR and increased positive behaviour scores. They suggest a mechanistic link with acupressure-associated opioid and serotonergic stimulation, and resultant bradycardia, and that it may be useful for calming horses in mildly stressful situations. However, it is not certain that a decrease in HR necessarily means that it is perceived as a positive reinforcer (Sankey et al. 2010a). Interestingly, the HR of horses that displayed stereotypic behaviour responded differently to human grooming (Normando et al. 2007).

Janczarek et al. (2018b) investigated the behavioural response of horses to stroking of different body regions (head, neck, trunk, forelimbs, and hindlimbs) with regard to sex, horse/pony type, and ‘emotional excitability’ (HRV) at rest and during a novel object test. In a random order, each site was stroked by hand (described as superficial, with mild, even pressure, using a relaxed hand, moving at a frequency of 25 moves per minute), for 5 minutes on each side of the horse (10 minutes total) by the same familiar experimenter. On average, stroking at all sites was perceived positively by the horses and ponies (behaviour scores > 3). Stroking of the head may be perceived more positively than the other areas. However, the trunk region included the withers and scapula, areas commonly involved in allogrooming, as well as areas that are not, such as the abdomen, ano-genital area, and loin, so region preference may have been confounded because commonly and uncommonly allogroomed sites were mixed in the one region.

Lansade et al. (2018) used grooming to induce emotions of contrasting valence in horses. They employed a putatively positive, individually responsive method (dynamically adapted during treatment according to handler's perception of the horse's response), and a putatively negative, fixed standardised method (based on previous observations of horse-rider dyads). The valence was verified by examining the frequency of approach and avoidance behaviours that the horses displayed during grooming, which differed significantly in the expected directions.

Sankey et al. (2010a) concluded that tactile stimulation of horses by humans is not perceived by horses to be as rewarding as food when used as a primary reinforcer in a learning task. They found a better and faster training outcome, decreased latency to approach, and increased time spent near the experimenter, after positive reinforcement training with carrots rather than with three wither scratches. However, the method of tactile stimulation may not have mimicked that of allogrooming or the previously mentioned studies.

Despite deficiencies in the design and reporting of some of these studies, they are very useful. They present information on mimicry of a naturally occurring behaviour that horses are strongly motivated to perform, which appears to be important for social cohesion and survival, and is likely to be positive emotion inducing. However, it is unclear which behaviour is pleasant, the performance or the receipt of grooming, or perhaps both. A study investigating social licking in cows found that receivers' HR decreased whilst performers HR increased (Laister et al. 2011), indicating a cost to the performer and a benefit to the receiver, which might support the notion that receiving grooming is the pleasant part.

Tactile stimulation may be forced or voluntary, and be a fixed/invariant procedure or variable/responsive to feedback from the individual animal. Methods vary by equipment used (e.g., hand, fingernails, brush/comb type), site/body area, duration, pressure, motion pattern, and stroke rate. Standardisation of the procedure may be difficult. Habituation to the procedure, personnel, and the test environment prior to testing is advisable (McBride et al. 2004). The use of control periods and treatment order should be considered carefully (Normando et al. 2003).

Behaviour indicative of positive perception of tactile stimulation may include approach, investigation, and voluntary engagement. Movement frequency prior to, during, and after the procedure was suggested to indicate positive anticipation, relaxation, and a desire for brushing to continue (Tamioso et al. 2018). In contrast, an animal that displays active avoidance, moves away, becomes engaged in other activity, or displays aggressive behaviour may have a negative perception of it (Hall et al. 2018). Care may be required to distinguish whether the behaviours indicate a neutral, or non-positive, versus a negaitve perception.

3.4.3 Food

In horses, taste preferences may vary by breed and sex (Janczarek et al. 2018a). The neurohormonal control of food intake is extremely complex, involving homeostasis, differing motivations for eating, and differing reward mechanisms. Additionally, understanding of satiety is incomplete (Esch et al. 2006; Smith et al. 2008). Whilst food has been used in many studies as a positive reinforcer for learning, or to induce a presumed positive association or positive emotion, the complexities of the neurophysiological systems involved with food may be oversimplified in the conclusions drawn. As a positive stimulus, food should be used with caution in studies on positive emotion. The behavioural and physiological responses of horses may be altered by anticipation (excitement, increase in HR, head shaking, pawing) (Innes et al. 2008), ingestion (involves head and neck movement, facial expressions may be obscured or altered by eating), and digestion (altered autonomic and cardiac activity, neuroendocrine changes, endorphin release), which may or may not reflect changes in emotion.

3.4.4 Anticipation and expectation

It may be possible to manipulate anticipation and expectation to induce emotion. Wanting, or positive anticipation may be more rewarding to some than consummation of the reward itself. Dopamine mediated ‘wanting’ (motivation for future seeking or appetitive behaviour) is distinct from opioid mediated ‘liking’. Liking precedes wanting temporally, that is, you have to first like something to know that you want it in future, and physiologically (opioids control dopamine release) (Burgdorf et al. 2006).

Expectation (positive or negative) may impact the future reward value of an experience, and thus the emotional valence associated with it (Keeling 2019). Manipulation of expectation can induce a positive (expectation met or exceeded) or negative (expectation underwhelmed) experience. In positive contrast tests, animals receiving a bigger than expected food reward display behaviour suggestive of a more intense positive emotion. Similarly, where the outcome is better than the negative expectation, a more positive emotion may result.

The same type of stimulus can be used across all treatments to create either opposing emotional valence, or a gradient of valences. For example, in sheep, variation in feed type i.e., standard, enriched, or unpalatable (wooden pellets), was used to meet, exceed, or underwhelm expectation respectively, in the anticipatory and consummatory phases (Reefmann et al. 2009b, c). In dogs, anticipation of delivery of a food reward was used to induce a state of positive anticipation, and withholding of the food delivery to induce frustration (unmet expectation) (Bremhorst et al. 2019).

3.4.5 Olfactory

Anecdotally, horses appear to react negatively to odours such as alcohol, or methylated spirits, used for skin disinfection prior to injections. Although that reaction may be due to a negative association of the smell with the pain of injection, it may be possible to explore scent for potential to induce positive and negative emotion, either primarily, or via association.

The effects of a synthetic equine appeasing pheromone (naturally produced by lactating mares) have been tested in a study on fear responses to novelty in horses (Falewee et al. 2006). The experiment and results are controversial (Dodman et al. 2008; Pageat 2008). Further research on the pheromone as emotion inducing or modulating is required.

3.4.6 Auditory enrichment

The calming effects of music (auditory enrichment) have been investigated in several species, including cats, dogs, and horses (Houpt et al. 2000; Wells 2009; Bowman et al. 2015; Snowdon et al. 2015; Stachurska et al. 2015; Stachurska et al. 2017; Wiśniewska et al. 2019).

Heart rate and heart rate variability, response latency, handling, and other behaviour have been found to respond positively to different music types. Interestingly, Stachurska et al. (2015) also found a positive relationship with music and racing performance in horses. The effects of music may be short-term, and may vary with age, gender, and music type. It is also possible that music has a calming effect on humans, which in turn impacts animals.

3.4.7 Social separation or reunion

Briefer et al. (2015b) used social separation and social reunion in horses to induce presumed negative and positive emotion, respectively. Locomotory behaviour (walking, trot, canter, turning) while in paddocks or box stalls was shown to affect physiological variables and may have confounded results. Emotional valence was presumed based on inference of behaviour and knowledge of the strong motivation to reunite with a conspecific in gregarious species. The two social situations are dependent on each other. It is not possible to reunite an animal without first isolating them, and vice versa. It is accepted that separation of social species is likely to induce stress and negative emotions such as fear. However, it is unknown if relief from a negative situation produces positive emotion in the same way as a positive emotion inducing experience that is independent of experience of a negative one. Is it an absolute positive or just less negative than being separated?

A control situation involving no manipulation of the horses was included in the study as a ‘neutral’ valence emotional experience to create three levels of valence (Briefer et al. 2015b). However, questions remain over the theoretical neutral valence of emotion. Is a neutral emotional response possible to define, or is something simply more or less positive than the comparator? Does an animal’s emotional response always move on a continuum from negative, through neutral, to positive? The nature of the stimuli used in a study, co-dependence between stimuli, and potential effects on the variables should be considered carefully.

3.4.8 Play

Play has been defined as behaviour that has no apparent immediate function (Ahloy-Dallaire et al. 2018). It is more commonly seen in juveniles and rarely in horses over 3 years old. The occurrence of play behaviour is often touted as a sign of good welfare, with a decrease in play behaviour commonly associated with stressful situations (Fraser 2008).

Provision of opportunities for play could be considered for potential to induce positive emotion. However, the uncertainty surrounding its function and meaning with regards to welfare (Hausberger et al. 2012; Ahloy-Dallaire et al. 2018), as well as the rarity in adult horses, may limit its application as a positive emotion inducing stimulus in the research setting.

3.4.9 Warm-water immersion

Spring water immersion is used as a treatment to aid healing in racehorses with musculoskeletal injuries. In a study by Kato et al. (2003), subjects were immersed in warm water (38^{o}C) up to the level of the olecranon, with warm water also showered over their backs. Parasympathetic nervous activity, as measured by HRV, increased, despite no change in HR. Behavioural measures were not included in the study; however, the authors made a subjective statement that the horses appeared calm, comfortable, and relaxed. Based on human studies of foot bathing, they attribute warm water immersion as relaxing to horses.

3.5 Assessment of positive emotion

Due to the paucity of explicit research on positive emotion in horses, this review expands to discuss measures of positive emotion in other species. It is worth considering indicators of good welfare for utility as indicators of positive emotion. Good welfare and positive emotions are linked. An animal that has good welfare must be in a positive emotional state and experiencing positive emotion often (Keeling 2019; New Zealand Thoroughbred Racing 2019). However, an animal that is experiencing a positive emotion may not necessarily have good overall welfare. Some purported indicators of good welfare may also be indicators of positive emotion or positive emotional state.

The use of advanced technology to help measure and interpret physiological and behavioural responses will continue to allow advancements in the field of animal emotion research. Neethirajan et al. (2021) present a review of applicable biosensors, wearable and environmental sensor technology, and data processing, that are available currently or in the future. Sensors are available for continuous, non-invasive, real-time measurement of sweat and skin conductance, HRV, RR, temperature via infrared thermal imaging, pH, position in three-dimensional space, and detecting viruses and bacteria, as well as a range of environmental conditions. Kinematic methods allow for detection of movement of body landmarks (Guesgen et al. 2017; Whittaker et al. 2019).

The development of technology to allow continuous, automated, non-invasive data collection without the influence of human presence and integration is exciting for research and practical applications. Technologies for facial recognition, GPS tracking, productivity, and quality assurance are in commercial development for use in cattle and pigs. Future developments include facial expression recognition (as used in humans), drone cameras for recording behaviour and facial expression, wearable electroencephalography (EEG), electromyogram (EMG, for superficial muscle tension), olfactory and chemical sensors, and vocalisation analysis. Future technologies may include handheld sensors for measuring yet to be discovered biomarkers of emotion remotely through hair, skin, or discharges (e.g., nasal, lacrimal, vaginal). Ongoing challenges include improving accuracy, feasibility, and affordability, reducing artefact, integration of multiple sensors, and refining algorithms using machine learning for data processing and analysis. Neethirajan et al. (2021) highlight the importance of gathering valid baseline data and establishing normal ranges with which to make comparisons.

Changes due to positive emotional responses are generally considered more subtle and variable, and thus present a challenge for scientific study (Boissy et al. 2007b). So far, no single measure has been identified as a reliable, valid indicator of positive emotional valence. Work on biosensor technology, automation of data capture, including facial expression recognition, and algorithms and artificial intelligence for data processing, may allow advances in this field of research with time efficiency, simultaneous measurement of multiple factors, standardisation across studies, integration of results, and detection of more subtle changes (Neethirajan et al. 2021). Studies included below are highly relevant, influential, or novel. They are presented under the principal measure used.

3.5.1 Physiological measures

Many physiological measures investigated in the study of emotion in the past have been found to indicate arousal rather than valence (Kremer et al. 2020). More recently, some physiological measures have been linked with emotional valence. However, further research is required (Lansade et al. 2018).

Assessment of the autonomic nervous system (ANS) may yield information about emotional experience. The two branches of the autonomic system, the sympathetic (SNS) and the parasympathetic (PNS), are in a balance, which alters with changes in physical (exercise or training effects), pathological (disease or drug effects), and psychological states. The relationship between the two branches is complex. Broadly, the SNS predominates during stress and the PNS predominates during relaxation. Many studies have looked at aspects of the autonomic nervous system (ANS) and the relative input of sympathetic and parasympathetic (vagal) control as indicators of negative emotional valence, but few studies pertain to positive emotion. Markers of the autonomic nervous system are considered under the categories below.

Cardiovascular measures

Heart rate

Cardiac activity is controlled by the ANS, with HR representing the net effect of the sympathetic and parasympathetic branches. Increased SNS or diminished PNS activity results in acceleration of HR. Conversely, low SNS activity or high PNS activity causes deceleration of HR (Kowalik et al. 2017). Many factors may influence HR, including movement or physical exercise, and ill health, as well as psychological stress.

Physical activity can confound measures of cardiac and respiratory function. Jansen et al. (2009) attempted to separate increases in HR due to emotion from that attributable to physical activity using a mathematical model. The experiment used a novelty (i.e., fear) test whilst a horse was exercised, which resulted in increased HR. The method may not be applicable for situations where HR decreases e.g., due to relaxing stimuli or low physical activity, but it may be useful for high arousal positive emotions e.g., anticipation of feeding, mating, or social excitement.

Heart rate has been used to determine emotional arousal, rather than valence, upon which other potential indicators of emotion were assessed (Briefer et al. 2015b; Briefer et al. 2017). However, in Briefer et al. (2015b) emotional valence was presumed from context (social separation and reunion), with a no manipulation control situation designated as neutral valence, both of which are assumptions. A higher emotional arousal level (2) was assigned to one treatment (‘All Leave’), despite the HR not differing from a treatment in the level below (‘All Return’). On that unsteady basis, RR was found to increase with arousal level. However, locomotion also increased with arousal level, which may have been a confounding factor for HR/arousal level, RR, and skin temperature. Locomotion was included as a fixed factor in the statistical models for heart rate variability, respiratory rate, and skin temperature, but not for HR. The order of treatments also had a strongly significant effect on RR and locomotion, as well as HRV, skin temperature, and all the other behavioural parameters, possibly indicating an effect of sensitisation or habituation to the stimuli. The authors do not seem to distinguish between physiological arousal resulting from exercise and psychological arousal resulting from emotions. Circular arguments are apparent in the authors’ discussion of behavioural and physiological results. This calls into question the validity of their conclusions.

Heart rate, or inter-beat-interval (IBI, R-R, or NN), is likely to reflect arousal rather than valence. However, Brosschot & Thayer (2003) found that HR elevations in humans are more prolonged with negative emotions than with positive emotions. Thus, the duration of HR elevation may be a useful factor to consider as a potential indicator of positive emotion.

In horses, the median time taken (latency) to reach maximum HR was longer in a putatively positive stimulus condition than a negative one (images of happy and angry human facial expression) (Smith et al. 2016). However, absolute maximum HR, average HR change from baseline, and HR recovery time did not differ.

A double-blinded placebo-controlled study investigating the effect of a synthetic equine appeasing pheromone on the fear response of horses when walked through a novel fringed curtain measured HR and behavioural responses. The area under the curve (AUC) of the HR over time graph was used as the main measure of total effect on HR, and is suggested by the authors to represent the consequences of stress (Falewee et al. 2006). They also suggest that in future it would be useful to measure HR recovery following the negative emotion inducing event, and record baseline HR to allow individual horse reference (due to the high variability in HR in horses).

Heart rate variability

In humans, different emotions can be distinguished based on HRV independent of HR (Gehrke et al. 2011). HRV may be an important marker of emotional regulation ability in humans.

Variation in the time between consecutive heart beats on an electrocardiogram (ECG) is known as HRV. This variability is mainly under the control of the ANS. Analysis of the variation in time between heart beats is a non-invasive measure which yields information about the balance of sympathetic and parasympathetic nervous tone. Heart rate variability is affected by physical exercise, physical health, emotion, and mental health. Analysis of heart rate variability is a widely used method to assess fluctuations in ANS activity. It is used to assess sympatho-vagal regulation of cardiac activity with changes in physical (exercise or training effects), pathological (disease or drug effects), and psychological states or load (Boissy et al. 2007b; von Borell et al. 2007).

The sympathetic branch (sympathetic nervous system, SNS) of the ANS predominates during stress and results in increased HR and/or decreased HRV via the neurotransmitters noradrenaline (norepinephrine) and adrenaline (epinephrine). The parasympathetic branch (parasympathetic nervous system, PNS) of the ANS predominates in the relaxed state with effects mediated by acetylcholine released from the vagus nerve. The assumption is made that PNS predomination in animals is positive, as it has been linked to social and psychological well-being in humans (von Borell et al. 2007; Kok et al. 2010). Additionally, low vagal tone may mean that an individual is more vulnerable to stress, as vagal tone is an indicator for physiological and psychological flexibility of an organism's stress coping ability. It has been suggested that HRV can indicate emotional arousal as well as emotional valence in animals, with measures of SNS indicating arousal, and measures of PNS indicating valence (Zebunke et al. 2011; Kremer et al. 2020).

Von Borell et al. (2007) provide a comprehensive review of the use of HRV measures in a variety of animal species. Although there can be large inter-individual differences, HRV has been shown in horses to have good stability for individuals across age and time. Genotype, behaviour, environment, temperament, and nutritional status are important in the large inter-individual variations in basal HRV of horses (Gehrke et al. 2011). There may be a gender difference, with females having higher PNS tone, but this was not conclusive. Time of day should be controlled for in experiments, as some indices change at night-time, with the PNS predominating in the resting horse.

Heart rate variability analysis is suggested to be a better indicator of physiological stress responses in horses than measures of cortisol, adrenaline, and noradrenaline, which are invasive, harder to measure, and are not always correlated with behavioural signs (Rietmann et al. 2004a). Psychological states can impact HRV without any detectable change in HR or RR (von Borell et al. 2007). Heart rate variability has been used in studies on stress, pain and analgesia, exercise and training, and temperament and coping in horses.

Measurement and analysis of heart rate variability

The measurement, calculation, and interpretation of HRV is complicated. Methods of measurement of IBI and the calculation of other HRV parameters are not universally equivalent and are prone to artefact. In addition, it can be difficult to separate the influence of motor (physical exercise), non-motor (emotional response), or disease effects on HRV. It is suggested that, for comparison of non-motor components of cardiac activity, only measures made during times of similar behaviour patterns should be compared (von Borell et al. 2007). This may be practically difficult as experimental treatments may induce different behavioural reactions, plus anticipatory and post treatment effects.

The time lapse in milliseconds (ms) between one R wave peak and the next is known as the R-R interval, or Inter-beat Interval (IBI, NN). The IBI is irregularly variable in normal, complex systems. It is the raw measurement from which all other indices of HRV are calculated. The instantaneous heart rate is directly related to the IBI (HR/minute = 60,000/IBI in ms), thus reporting of mean IBI adds little further information than mean HR. Measures of the variance of IBI give more information.

Within the category of HRV there are many different indices that can be calculated and analysed. The types of indices include time domain, frequency domain, geometric, and non-linear (e.g., Poincare plot) approaches. Time domain and frequency domain methods are more common in the literature.

Time domain measures include indices such as SDNN (standard deviation of normal-to-normal IBI, also known as SDRR), which indicates short and longer term variability from sympathetic and/or vagal activation (overall variability), RMSSD (root mean square of successive differences of all IBI), which indicates short term variability from vagal activation (higher values indicate increased parasympathetic activity), RMSSD/SDNN ratio, which reflects the balance between SNS and PNS, and PNN50 which is the proportion of pairs of successive IBIs that differ by more than 50ms divided by total number of IBIs. Generally, a decrease in SDNN or RMSSD:SDNN ratio reflects a predominance of the SNS and decreases with increasing stress, whereas an increase in RMSSD or RMSSD:SDNN ratio reflects a predominance of the PNS, and might be expected to increase when relaxed or experiencing pleasure (Visser et al. 2002; Boissy et al. 2007b; von Borell et al. 2007; Zebunke et al. 2011; Wang et al. 2012).

The PNS is associated with rhythmic pulses at a higher frequency than SNS. This gives rise to frequency domain measures of HRV. Power spectral density analysis (PSD or PSA) of HRV shows two major peaks in frequency. The high-frequency (HF) peak is due to respiratory sinus arrhythmia from PNS activity. The low-frequency (LF) peak reflects mainly SNS, and to a lesser extent, PNS activation. The very low frequency (VLF) band indicates sympatho-vagal balance in humans (Usui et al. 2017). Analysis is commonly performed using fast Fourier transformation (FFT), or less commonly autoregressive PSA, on data from blocks of 256, 348 or 512 consecutive beats, to give LF band, HF band and total power components, and a ratio of LF/HF. LF and HF should be expressed in normalised units, i.e., as a percentage of total power. The band frequencies differ with species, and HF band location is affected by respiration rate (Kuwahara et al. 1996; Rietmann et al. 2004a; von Borell et al. 2007).

Indicators of parasympathetic nervous system activity include HF, HF/total power (PNSI), and RMSSD. Indicators of sympathetic nervous system activity include LF, LF/HF (SNSI), and VLF.

The most commonly reported HRV measures in the literature are mean IBI, RMSSD, SDNN, RMSSD/SDNN ratio, HF, LF, and LF/HF ratio. It is suggested that time domain measures be used for analysis of short HRV recordings, as analysis of HF requires a minimum of 1 minute recording, and LF a minimum of 4 minutes (10 times the lower bound of the frequency band) (von Borell et al. 2007).

Equipment used to measure the IBI, from which HR and HRV indices are calculated, may be a heart rate monitor (HRM) or an ECG recorder. An HRM consists of a two electrode flexible/textile elastic belt with onboard memory and/or a transmitter, for example Polar Equine RS800CX (Polar Oy, Finland) (Lansade et al. 2018). Heart rate monitors detect electrical impulses, associating the largest peak change in voltage with the R wave of a heartbeat (ventricular depolarisation), and records the time interval between them. In horses, the T wave of a heartbeat is very pronounced and can be mistaken for an R wave (von Borell et al. 2007). An ECG detects the electrical impulses from all parts of the heartbeat (PQRST waves) and records the voltage change over time. The IBI data is then calculated from these ECG traces. ECG recordings may be more accurate and have fewer artefacts than those from an HRM. Agreement between HRV data from an HRM and an ECG for stationery horses may be acceptable. However, all HRV indices (RMSSD, SDNN, LF, HF, and Poincare plot) except mean HR and mean RR interval, have poor agreement for moving or exercising horses (Lenoir et al. 2017). Although more complex to use, an ambulatory ECG monitor, e.g. Holter monitor (Del Mar Avionics), may be advisable to allow more accurate measurement of IBI, particularly in moving horses (Gehrke et al. 2011).

Care needs to be taken with the methods of post-processing used to produce the final data for analysis. Correction of artefacts and data processing (e.g. handling of outliers, missing data, interpolation, or smoothing of data) can have a large impact on the data used for HRV analysis, and affects the comparison of results between studies (von Borell et al. 2007). As well the high amplitude and variation of the T wave in horses (causing artefacts in IBI a few milliseconds apart), ectopic beats due to high PNS tone (IBI's show large differences) and physiological second degree arterio-ventricular (AV) block (IBI twice the previous) are common in horses. Artefacts may also occur due to muscle contractions and movement of electrodes on skin. Fully automated algorithm functions included in the software of measuring devices may not be accurate in horses (Lanata et al. 2015). However, visual inspection and manual manipulation of data may potentially introduce bias. Some proprietary software (e.g., Polar ProTrainer Equine Edition) may not provide details on data algorithms due to commercial sensitivity, which does not provide the transparency necessary for robust research. Alternative algorithms for removal of movement artefact through combined use of a triaxial accelerometer have been investigated with good results, which may enable ECG measurement in ambulatory conditions and non-linear analysis (as is the case in human ECG analysis) (Lanata et al. 2015). IBI data containing more than 5% anomalies, or segments containing three or more consecutive errors, should not be included in analysis (von Borell et al. 2007). Splicing of segments of data together is not recommended for frequency domain measures, as it interrupts the underlying time series which analysis is based on. The recording device and associated post-processing software should be chosen carefully. The type of equipment used, how it was used, the method of data checking and editing, and any software settings used, should be stated when reporting results.

Calculation and analysis of the HRV indices is performed using various software (e.g., Kubios HRV software, University of Eastern Finland), details of which should also be published. Calculation of time domain measures (e.g. RMSSD, SDNN) is fairly straightforward, however spectral power analysis for frequency domain indices (e.g. HF, LF) is more complicated to calculate and interpret (see von Borell et al. 2007 for further information). In particular, there is variation in use of species appropriate frequency band widths and the corresponding respiratory rate, and stationary conditions (for examples see Kuwahara et al. 1996; Kato et al. 2003; von Borell et al. 2007; Gehrke et al. 2011).

Research involving HRV and emotion

In humans, RMSSD is lower in those with depression, and SDNN increases with pleasant tactile stimulation (Kremer et al. 2020). Kok & Fredrickson (2010) found that increases in parasympathetic tone (as measured by respiratory sinus arrhythmia) in humans predicted increased positive emotions and social connectedness, and vice versa. It is suggested that parasympathetic tone may be a useful indicator of psychological and physiological flexibility, and useful for assessing ability to cope with stress, with low vagal (PNS) tone associated with more vulnerability to stress (von Borell et al. 2007; Kok et al. 2010).

An effect of treatment (familiar human presence or brushing by a human) and genetic line (more/less reactive to social isolation) on HR, RMSSD/SDNN ratio, and LF/HF ratio was found in sheep (Tamioso et al. 2018). However, the findings for time domain (RMSSD/SDNN) and frequency domain (LF/HF) measures showed differing patterns of response. Further research is needed on autonomic responses of animals with different temperament traits.

In sheep, RMSSD increased with grooming (putatively positive emotion) (Reefmann et al. 2009a; Coulon et al. 2015). In sheep and dogs, RMSSD decreased in response to social isolation (putatively negative emotion) (Reefmann et al. 2009a; Kremer et al. 2020). In pigs given a food reward following an operant learning task, authors observed increased arousal (increased SDNN, decreased RMSSD/SDNN ratio initially, prolonged HR elevation) caused by reward anticipation which was positively valenced (increased RMSSD), and concluded that pigs were experiencing a positive emotion during feeding (increased RMSSD/SDNN ratio) (Zebunke et al. 2011). However, when arousal was controlled for in goats, RMSSD was indicative of increased arousal rather than valence (Briefer et al. 2015a).

Heart rate and HRV have been used as measures of relaxation or excitability in geriatric horses exposed to 'new age relaxation' music with a 2-3 week relaxing effect found (Wiśniewska et al. 2019). However, no effect on HR or HRV was found in horses that were groomed either pleasantly or aversively (Lansade et al. 2018). The study design was parallel rather than crossover, and it is not clear if individual variation in baseline (either prior to the first session or before each subsequent session), which can be high in horses, was accounted for. There was a tendency towards a lower SDNN with the aversive treatment. However, the RMSSD/SDNN ratio was not calculated/analysed. The authors suggested that, contrary to other studies on grooming/tactile stimulation in horses where horses were relaxed, horses in their study were aroused in both the positive and negative treatment as evidenced by the active avoidance or approach behaviours seen. They interpreted the results to mean that both groups had similar emotional arousal levels.

Kowalik et al. (2017) reported that racehorses exposed to 30-minute massage sessions by a person three times per week over 7 months showed increased RMSSD, decreased LF, increased HF, decreased LF/HF ratio at rest, during saddling, and during warm up walking. Racing performance also improved (number of races won, winnings per race, and success co-efficient).

Janczarek et al. (2018b) used HR and HRV (RMSSD, LF/HF ratio) as a measure of 'emotional excitability' in horses at rest, during a novel object test, and in response to stroking various regions of the body. While the study findings are limited, the results may serve as a useful comparator for other research. Interestingly, there were no differences between the HR and HRV results from stroking the left and right sides of the horse, despite horses being more often handled on their left side.

There are many indices of HRV, some of which are thought to relate to emotion. Calculation and interpretation are not always straightforward, and arousal level may confound. Although complex to measure, calculate, and interpret, HRV warrants further investigation as a potential indicator of emotional valence and/or arousal.

Respiratory rate

Despite RR decreasing with a positive stimulus (better than expected food reward) and increasing with a negative stimulus (worse than expected food reward, wooden pellets) in sheep (Reefmann et al. 2009c), in goats it was linked to arousal rather than valence (Briefer et al. 2015a).

Peripheral temperature

The term emotional fever refers to the phenomenon where a typically negative emotional experience causes a drop in peripheral temperature due to peripheral vasoconstriction, and an increase in core temperature due to diversion of blood to vital organs (Kremer et al. 2020).

Core and peripheral temperatures fluctuate due to inflammation, normal cyclic patterns, and reproductive status. However, activation of the hypothalamic-pituitary-adrenal (HPA) axis and ANS with physiological or psychological stress, and subsequent catecholamine and cortisol increase, can also result in heat production and vascular changes, leading to alterations in temperature (Church et al. 2009). Changes in the ANS result in changes in the diameter of blood vessels (broadly, vasoconstriction with SNS activation and vasodilation with PNS activation) (Church et al. 2009). Radiated temperature is more responsive than core temperature (Church et al. 2009). It is suggested that there are species specific differences in sympathetic vasoconstriction of specific body regions (e.g. eye, nasal planum) in response to acute stress or fear (Stewart et al. 2008a).

Infrared thermography (IRT) cameras can be used to collect real-time images (radiometric JPEGs) of the radiated electromagnetic energy at a distance from the subject, sometimes without need for contact and/or restraint, making it non-invasive and non-contact. Temperature data is embedded in each pixel (sensor), effectively mapping the two-dimensional temperature of the object, allowing for measurement of temperature at different sites over time. Another appeal of IRT technology is the potential for automation of image capture, which has been developed for use in cattle, along with the use of electronic identification, in early disease detection (Stewart et al. 2005).

Although IRT equipment is portable and simple to use, thermogram image capture and analysis is complicated. It is influenced by the subject, environment, and image acquisition and processing (Church et al. 2009; Stewart et al. 2010a; Rekant et al. 2016). Thermograms of healthy animals can vary cyclically and/or by species. Hair may affect radiated heat, which may be relevant when measuring areas of differing hair density or coarseness, shaved or clipped coats, or the margins between haired and hairless areas. Skin colour also affects solar loading and radiated heat; black skin is warmer than white areas during the day but cooler at night. Dirt, moisture, and foreign material on the targeted body part may impact the thermogram. Age affects radiated heat from the eye of humans and may also affect other sites. Inflammation may alter temperature, so it is important to ascertain that the area of interest is disease free. Some drugs, in particular sedatives, analgesics, and anaesthetics, can impact core and peripheral temperatures.

Environmental factors include ambient temperature and relative humidity, wind or drafts, direct sunlight, rain, and other weather conditions. The variation in weather conditions, which can affect the temperature of the body part as well as affecting image acquisition, may overshadow the temperature differences in the subject (Rekant et al. 2016). The solar reflectivity of nearby surfaces may also impact measurements (Dai et al. 2015).

The specifications of the infrared camera have a direct impact on results. Infrared cameras differ in infrared sensor type, sensor resolution (as distinct from screen or video image resolution), sensitivity (minimum detectable change e.g., 0.02-0.075^{o}C), accuracy (maximum margin of error e.g., 2% or ±2^{o}C), exposure time, and frame rate (FLIR Systems 2016). Spot size (the area covered by each pixel on the target), and field of view also impact accuracy.

The distance between the subject and the camera position may affect results. Some cameras have settings or algorithms to account for distance, emissivity, ambient temperature, and relative humidity. Re-calibration may be required regularly during use to compensate for ambient temperature. Cameras need to warm up and equilibrate to the environmental temperature before use. They should be kept away from heat sources and out of direct sunlight. Some cameras automatically correct for drift or non-uniformity correction (NUC) periodically during use. The various software required for processing of images and data during live viewing or post capture can also impact the data.

Skin temperature and humidity

The ANS controls blood distribution to the skin, which in turn impacts the skin surface temperature (Hall et al. 2018). In sheep, the skin temperature over the withers and nasal temperature increased after brushing (putatively positive emotion inducing) (Kremer et al. 2020). The variability of body-surface humidity indicated emotional valence in sheep, but body-surface temperature did not (Reefmann et al. 2009c; Reefmann et al. 2009a).

The nasal skin temperature of cows decreased by 0.36°C when experiencing a putatively low arousal positive emotion induced by human stroking of a preferred region (Proctor et al. 2015). That contrasts with studies finding that peripheral temperatures decrease with high arousal negative emotions such as stress, fear, and frustration. However, temperature was measured on the central exterior part of the nose, i.e., in the middle of the nasolabial plate, which, in cattle, is kept moist by underlying glands (Proctor et al. 2015). Moisture may affect temperature through evaporative cooling, but was not assessed in that study, which may explain the contrary results.

Skin temperature has been found to vary with emotional arousal and valence in horses but is inconsistent (Briefer et al. 2015b). Skin temperature was affected by locomotion, ambient temperature, age, and weight (Briefer et al. 2015b). Species, context, reproductive status, time span, and ambient temperature and humidity may be important factors in measuring skin temperature and humidity.

Eye temperature

Eye temperature (ET) has been used as a non-invasive measure of pain, fear, and stress in cattle and other species (Stewart et al. 2005; Stewart et al. 2008b; Stewart et al. 2008a; Stewart et al. 2010a). The maximum temperature recorded at the medial canthus of the eye, measured using infra-red thermography, decreased with pain (disbudding, electric prod), fear (aversive handling, electric prod) and stress in cattle. After an initial rapid drop in ET during the five minutes following disbudding without local anaesthetic, ET increased above the baseline during the rest of the 40-minute period (Stewart et al. 2008a). The ET increase was not thought to be a result of inflammation, hot iron cautery, increased physical activity, or increased HPA activity. Administration of local anaesthetic alone resulted in an increase in

ET. Although administration of adrenocorticotropic hormone (ACTH) did not result in a change in ET, it was shown to decrease rapidly with infusion of epinephrine/adrenaline in cattle. This supports the theory that SNS activation, rather than the HPA axis, mediates the rapid initial eye temperature decrease with acute negative emotion inducing events, possibly via vasoconstriction (Stewart et al. 2008a; Stewart et al. 2010a).

Contrasting or inconsistent results have been seen in horses (Cook et al. 2001), deer (Cook 2005; Cook et al. 2006), sheep (Cannas et al. 2018), and piglets (Yáñez-Pizaña et al. 2019). Eye temperature over the lachrymal gland in deer increased along with cortisol in response to velvet antler removal, which is a painful procedure (Cook 2005; Cook et al. 2006). The temporal relationship of the ET measurements to the procedure and the frequency of sampling varied. Perhaps the timing and frequency of measurements in relation to the negative events may account for why an acute initial decrease in ET was not seen in these studies, whereas it was seen in cattle undergoing similar procedures.

A study on the impact of environmental enrichment and maintenance of social grouping (putatively positive experiences) during weaning (putatively negative experience) of piglets found significant differences in eye, nasal, and pinna temperatures between groups (Yáñez-Pizaña et al. 2019). However, due to the inconsistent results, use of these parameters to assess stress reduction efforts was not supported.

In horses, measurement of ET has been used to assess fever, stress from competitions, the fit of different bridles, coat clipping and other fear inducing procedures, housing system, physical exercise/fitness, and poor performance (Jansson et al. 2021). In horses at rest in their home environment under field conditions, maximum ET at the medial canthus/lacrimal caruncle area ranged from 29.4 to 37.6^{o}C (Jansson et al. 2021). ET varied by breed, sex, individual, season, farm, environment, and operator, but not by eye side, age, or height (Jansson et al. 2021). There was no correlation with rectal temperature. The inter-individual variation was larger than the intra-individual variation, meaning that ET may still be useful for detecting differences, even in long-term field studies with repeated measures. It was suggested that changes in pupil size may account for the variation that was seen between indoor and outdoor recordings. The large farm effect may be due to differences in diet or surface emissivity, but the extent of variation means that direct comparison between farms is not recommended.

In horses, maximum ET was moderately positively corelated with salivary and plasma cortisol levels following intramuscular administration of ACTH in simulation of HPA axis stimulation (Cook et al. 2001). Samples were collected at 20-minute intervals to 3 hours 20 minutes after administration of ACTH. Thus, an acute initial temperature drop, such as that seen in calves, would not have been detected. In contrast, no corelation between serum cortisol and ET was found in a study on physiological stress in endurance horses (Redaelli et al. 2019).

In horses participating in a ridden jumping competition, a weak correlation (Spearman's rank correlation coefficient = -0.25) was found between horses that had a higher ET and poorer performance (Bartolomé et al. 2013). Although results were affected by genetics, and possibly individual temperament traits, the authors suggested that horses with a higher ET were more stressed.

In a study involving a small number of horses ($n = 5$) undergoing a putatively aversive procedure (fastening the noseband on a double bridle), an increase in ET over time was seen across all treatments, including the control situation, where the noseband was not fastened (McGreevy et al. 2012). Eye temperature did not vary with the tightness of noseband at 5- or 10-minutes post treatment. However, it was suggested that the increase in ET indicated a stress response to the double bridle itself, although the method and interpretation have been criticised, and the conclusions are not supported.

In an experiment testing fear response in horses, ET was found to increase by an average of 0.29^{o}C following a novel object test, with a tendency ($p < 0.1$) to differentiate between more and less fearful horses (Dai et al. 2015). The authors suggested that the time for ET to return to baseline levels should be investigated in future studies.

Many intrinsic and extrinsic factors can influence eye temperature and measurement (Church et al. 2009; Jansson et al. 2021). Some changes in peripheral temperature may be species specific (Stewart et al. 2008a). Methodological and analytical differences between studies can make comparison difficult. The temporal relationship of the timing and frequency of measuring eye temperature in relation to treatment may have a large impact on results.

Most of the cited studies involve situations of likely high arousal negatively valenced emotion. It may be that ET is an indicator of physiological stress or arousal rather than valence. However, it is theorised that ET might increase in response to positive emotion (Church et al.

2009), possibly through stimulation of the PNS. ET warrants further investigation as a non-invasive measure of positive and negative emotion, but caution is advised with interpretation of values, particularly those within the margin of error of the measuring equipment.

Other physiological parameters

Brain imaging (e.g. functional magnetic resonance imaging (fMRI) and positron emission tomography (PET)) and measures of neuronal activity (e.g. functional near infrared spectroscopy (fNIRS) and electroencephalography (EEG)) show promise for expanding knowledge of emotion processing in some species of animals (Burgdorf et al. 2006; Wager et al. 2008; Muehlemann et al. 2011; Berns et al. 2012; Gygax et al. 2015; Alvarez-Bolado et al. 2016; d'Ingeo et al. 2019; Stomp et al. 2021).

Changes in levels of various neurochemicals may be useful indicators of positive emotion. Direct measurement of central neurotransmitters or neuroendocrine metabolites in brain tissue or cerebrospinal fluid is invasive and difficult (Burgdorf et al. 2006). Correlation of central levels with peripheral measures is unconfirmed for many chemicals, with more investigation of measurement techniques, variation over time (circadian and other rhythms), and validation required. Research has been conducted in humans, horses, dogs, pigs, rats, and bumblebees on levels of dopamine, serotonin, oxytocin, endorphins, and opioids, and administration of dopamine agonists and antagonists (Burgdorf et al. 2006; Çakiroğlu et al. 2007; Kalbe et al. 2010; Alberghina et al. 2017; Lansade et al. 2018; Marcet Rius et al. 2018; Kremer et al. 2020; Lesimple 2020).

Blood cortisol did not differ between horses that were subjected to a positively or negatively perceived grooming experience, despite active avoidance and approach behaviours (Lansade et al. 2018). However, cortisol has been found in some situations to be lower after positive experiences in humans and horses (Lansade et al. 2018). It is suggested that glucocorticoid levels reflect emotional arousal rather than valence (Iyasere et al. 2022).

In humans, oxytocin is associated with feelings of well-being. It is unclear if oxytocin is related to momentary emotion, emotional state, or both. It is suggested that it might be an indicator of positive emotion and good welfare in animals (Odendaal et al. 2003; Mitsui et al. 2011; Macchitella et al. 2017; Petersson et al. 2017; Lansade et al. 2018; Keeling 2019). However, research on oxytocin in different species has provided inconsistent results (Rault et al. 2017;

Marcet Rius et al. 2018; Kremer et al. 2020). The validity and reliability of current measures of central and peripheral oxytocin has been questioned (Kremer et al. 2020). The role of oxytocin in emotional valence is unclear, and the premature use of oxytocin as an indicator of social, positive, or psychological well-being has been criticised (Rault et al. 2017).

In horses experiencing positive and negative tactile stimulation (grooming), plasma oxytocin did not change following treatment, but the decrease in basal oxytocin level was suggested to be a marker of improved general wellbeing with the pleasant grooming over time (Lansade et al. 2018). Higher basal oxytocin levels are seen in humans with greater relational distress (Lansade et al. 2018).

In humans, positive emotions and emotional states have been associated with cardiac disturbances and decreased inflammatory markers, such as fibrinogen, c-reactive protein, cytokines, and immunoglobulins (Steptoe et al. 2005). Markers of immune function might also be applicable indicators of animal emotion (Boissy et al. 2007b), although conflicting results have been found with salivary secretory immunoglobulin A (SIgA) in calves (Lv et al. 2018; Spiesberger et al. 2019).

Emotional valence can be communicated between humans and animals via chemosignals (Semin et al. 2019; Calvi et al. 2020). Chemosignals, or semiochemicals, may be classed as odours or pheromones, as distinct from volatile molecules (Calvi et al. 2020). It may be that future research in this field identifies species specific chemosignals denoting positive and negative emotion in animals.

Fields such as genomics, proteomics, and metabolomics may be useful in future research on emotion and emotional disorders (Kremer et al. 2020). MicroRNA expression, and other molecular and cellular changes, have been associated with environmental enrichment in healthy rats, as well as neurologically or psychologically dysfunctional ones (Kuznetsova et al. 2020).

Parameters, such allostatic load, telomere length, and gut biota, may represent the accumulation of negative and positive emotional states over time, rather than momentary emotion. They may be promising in animals, particularly as further knowledge develops about them in humans (Korte et al. 2007; Shammas 2011; Mauss et al. 2016; Ochoa-Repáraz et al.

2016; Garrett 2017; Koopman et al. 2017; Panduro et al. 2017; Webb et al. 2018; Kremer et al. 2020; Lee et al. 2020; Sabry et al. 2020; Tsai et al. 2020).

3.5.2 Behavioural measures

Behaviour is a product of genetics, previous experience, and environment (Overall 2013; Grandin et al. 2014; Jensen 2015; Whittaker et al. 2019). An animal's behaviour may provide an indication of its emotional experience. Brando et al. (2018) described behaviour as the ultimate phenotype, reflecting an animal's choices and an expression of its emotions.

Behavioural measures are likely to be important when investigating positive emotion in animals, either as an indicator of emotion, or as part of the process of verifying the valence of the putative emotion to enable assessment of other indicators (Whittaker et al. 2019). However, interpretation of behaviour is complex and requires a cautious approach to avoid bias. The same behaviour performed in different contexts may imply different underlying emotion or motivation. The interplay between arousal and valence is also likely to impact behaviour (Kremer et al. 2020). Behaviours may be species-specific and be impacted by many factors including breed/strain, age, sex, reproductive status, social status, and altered mental health. Some behaviours do not occur often or for long enough to enable useful analysis. The work by McDonnell et al. (1995), McDonnell et al. (2002), and McDonnell (2003) provides a foundational ethogram of social behaviour in horses.

Some behaviours, termed displacement behaviours, may be displayed in positive, negative, or ambiguous situations (Lesimple 2020). For example, yawning or vacuum chewing in horses, and purring in cats, is often considered to signify relaxation, but may also be triggered by stress and negative contexts. Thus, it is essential to consider the context in which the behaviour is seen.

Behaviour can be assessed in many ways. Behaviour can be assessed as 'whole animal behaviour', or concentrated on specific body parts or regions e.g., facial expressions (Kremer et al. 2020). The occurrence, frequency, and/or duration of specific behaviours may be measured. Behaviour can also be scored, which involves judgement and interpretation of the whole animal and the context in which behaviour occurs. Although it is subjective, behaviour scoring has been shown to correlate with more quantitative measures in horses in some circumstances (LeScolan et al. 1997; Minero et al. 2009).

While the use of a scoring method for measuring behaviour has appeal due to time efficiency and ease of amalgamating a variety of behaviours into a single numerical variable, a cautious approach is required. For example, Janczarek et al. (2018b) used a behaviour scoring system that involved interpretation of a horse as positive or negative on a scale of 1 to 5. The descriptors of each score included attitudinal attributes and individual behaviours that may be interpreted in alternate ways by others. No measures of inter- or intra-observer reliability were given. Additionally, the score results provided for little differentiation between treatments and factors. If individual behaviours had been more objectively measured (e.g., frequency of occurrence, duration) there may have been more opportunity for teasing apart subtle differences and/or then conducting analysis, such as principal component analysis, (PCA) to discover patterns of behaviour.

Measurement of behaviour is time and resource intensive. Future automation and standardisation of behaviour measurement and study design may also aid with objectivity of results and allow more useful comparison between studies. Fureix et al. (2011) discuss a method to measure body postures in horses using geometric morphometrics, which may have potential application in animal emotion studies and lead to future automation of assessment. Until then, it is possible for behaviour measurement to be influenced subjectively (Whittaker et al. 2019). To minimise observer related bias, behaviours of interest need to be precisely described and catalogued, observers trained, and reliability measured (Lesimple 2020). Establishing the validity of behavioural indicators of positive emotional valence is difficult, as there are no gold standard markers with which to correlate and establish sensitivity and specificity (Whittaker et al. 2019).

Body posture and movement

In humans, facial expressions and body posture (body language) are known to correlate with human physical and emotional state (Keeling 2019). In animals, caretakers and veterinarians use body posture to infer the experience of pain. Research has been conducted on some specific body parts, such as head, neck, and tail position or movement, as indicators of emotion in animals (Keeling 2019; Kremer et al. 2020). However, no body parts or positions are definitive, most are affected by arousal level, and are context dependent.

A study on horses separated from and reunited with conspecifics found that during the first ~20-30 seconds of treatment, locomotory behaviour, frequency of turns, and head movements increased with arousal level (as determined by HR), and the duration of non-masticatory chewing movements decreased (Briefer et al. 2015b). Turns, head movements, and nicker vocalisations were inconsistently impacted by presumed emotional valence. Although the HR parameter that was used to determine the arousal level was likely affected by locomotion, which was not controlled for during analysis, the authors suggested that locomotion was a good indicator of arousal. They also suggested that head high duration and chewing duration were good indicators of valence, based on assessment of fit of models for arousal or valence using corrected Akaike's information criterion (AIC_C). However, the ΔAIC_C was less than two for head high, thus the valence model may not be a significantly better model than arousal, even though the AIC_C weight (probability of support for the model) for the valence model was 0.71.

Vocalisation

Vocalisations may convey non-verbal emotional information and are recognised as a marker of current momentary emotion (Knutson et al. 2002). Although observed more often in a negative context, such as separation from conspecifics, pain, or the presence of threat, vocalisations are associated with purportedly pleasant situations, and may reflect positive emotion in some situations (Keeling 2019; Whittaker et al. 2019).

Vocalisation can be measured in terms of occurrence of simple measures of different vocalisation types, frequency and duration of vocalisations, and energy distribution. Methods from human linguistics research can also be applied to aid deeper understanding. These methods examine the fundamental frequencies produced by the vibration of the vocal cords and vocal tract filtering (Whittaker et al. 2019). Automated recording and analysis of vocalisations means that measurement becomes less time-consuming, more objective, and more feasible to collect data over time with which to indicate emotional state (affective balance), as well as momentary emotion. However, application to animals would first require data and algorithms to be developed and verified for the conditions of interest.

Vocalisations may reveal information about emotional valence through frequency, and emotional arousal through increased rate of occurrence, volume, duration, and variability of frequency as arousal increases (Webb et al. 2018; Whittaker et al. 2019). In silver foxes, calling rate and time spent vocalising were indicative of arousal, whereas call type and maximum frequency amplitude indicated valence (Gogoleva et al. 2010). Higher frequency vocalisations are associated with positive situations in dogs, cats, pigs, and rats, whereas some other species emit lower frequency vocalisations (Knutson et al. 2002; Burgdorf et al. 2006; Yamamuro et al. 2010; LaFollette et al. 2018; Whittaker et al. 2019). However, it may be that filter-related vocalisation parameters indicate positive and negative emotional valence across species (Briefer et al. 2019).

Horse whinnies may include social and emotional information (Lesimple 2020). Snorts are found more in positive situations and less in negative ones, and may be an indicator of emotional valence. Snorts have been associated with positive experiences in horses, rhinos, and tapir (Stomp et al. 2018; Whittaker et al. 2019; Stomp et al. 2020). Whinny duration may also be an indicator of valence (Whittaker et al. 2019). Familiarity between the sender and the receiver of vocalisations may have an impact (Briefer et al. 2017).

Caution with interpretation is advised (Stomp et al. 2020). Arousal and valence may both influence vocalisation. Context is important. Some vocalisations that are traditionally accepted as indicating positive emotion may also be emitted in other contexts (Keeling 2019). For example, cats may purr during food solicitation, and when anxious. Anticipation may influence vocalisations with ambiguous results (Whittaker et al. 2019).

Facial movement or expression

Emotion can be shown via facial expression (Langford et al. 2010). Observable changes in the appearance of the face due to movement, often termed facial expression, can be defined in terms of a single behaviour (e.g. blink), as a collection or pattern of behaviours, or by the underlying specific facial muscle contractions or relaxations (Waller et al. 2020). Facial expressions are characteristic of certain emotions in humans e.g. joy, anger, and sadness, and may be an emotional display in animals (Whittaker et al. 2019). There is evidence for homology and shared ancestral evolution of facial expression across species (Wathan et al. 2015; Whittaker et al. 2019; Waller et al. 2020). The facial expressions of rats, mice, and chimpanzees were found to be similar to humans in trials involving pleasant and unpleasant tastes (Paul et al. 2005).

Whilst research on facial expression has mostly investigated pain, a negative emotion, it has enabled facial expression to be regarded as a valid scientific measure in animals. Facial expression is being explored as a possible indicator of positive emotion in animals. Generalised facial relaxation may indicate positive emotion, with play faces generally described as relaxed and open mouthed (Descovich et al. 2017).

Facial expressions and vocalisations may reflect a need to communicate emotion, and thus be strongly influenced by the 'audience' (Paul et al. 2005). Communication via facial expression may have adaptive value in warning of threat or danger, seeking care, or providing positive feedback (e.g. allogrooming, play), for example (Whittaker et al. 2019). Horses exposed to videos of a human interacting with another horse in a positive and a negative way displayed differences in facial expressions, HR, and contact seeking behaviour (Trösch et al. 2020). The authors suggest that horses can perceive the emotional valence of third-party intraspecific encounters, lending further weight to the communicative function of emotional displays. The display of facial expressions may be dependent on the presence of a suitable audience and may vary with the degree of visible social communication in a species. To allow variation in facial expression, the underlying structures and musculature must be sufficiently complex, and visual capability adequate to detect facial changes. The prey or predatory nature of a species, and the context, may also influence the display of emotions that suggest vulnerability. For example, it is suggested that rats may conceal signs of weakness with pain until it is intense (Lezama-García et al. 2019; van Dierendonck et al. 2020). The potential suppression of visual signs of emotion adds to the challenge of researching indicators of emotion (Hall et al. 2018). However, facial expressions may be a more accurate measure of emotion than gross body movements, and are less able to be suppressed than other indicators in humans (Descovich et al. 2017).

Along with a communicative function of facial expression of emotion, expressions may serve to regulate sensory exposure of the eyes, nostrils/nares, and mouth (Susskind et al. 2008). For example, wrinkling of the nose and eyes with disgust when an unpleasant smell is detected to reduce sensory exposure, or widening the eyes with fear to increase sensory exposure. It is also suggested that facial expressions convey signs of intent (e.g., to attack or withdraw) as well as emotion (Lezama-García et al. 2019). It is not clear if facial expression in animals can induce a hedonic state via facial feedback like a smile can in humans (Whittaker et al. 2019).

Facial expressions can be systematically identified and described using facial action coding systems (FACS) and grimace scales, as well as other specific ethograms. Facial expression in animals is utilised for pain recognition and grading in many species, including mice, rats, sheep, cattle, pigs, cats, rabbits, and horses, via validated grimace scales (Flecknell 2010; Langford et al. 2010; Leach et al. 2011; Sotocinal et al. 2011; Dalla Costa et al. 2014; Keeling 2019; Whittaker et al. 2019; Mogil et al. 2020). Originating from pain scoring in verbal and non-verbal humans, facial grimace scales for pain measurement have been developed in ten animal species (Mogil et al. 2020). In horses, signs of pain are identified by ear position, tension above the eyes, jaw clenching, and the appearance of the nares and chin, along with changes in composite pain scores and a decrease in alertness and exploratory behaviour (Dalla Costa et al. 2014; Lezama-García et al. 2019). Automated assessment of pain using computer image analysis has been developed in mice with high accuracy (93%) compared to human coders (Mogil et al. 2020) and is in development for use in horses (Andersen et al. 2021). However, the accuracy and validity of facial expressions in assessment of pain and welfare in horses has been questioned and further research is required (Lesimple 2020). For a review of grimace scales in pain measurement, including confounding factors, see Mogil et al. (2020).

The human Facial Action Coding System (FACS) based on facial muscle movements was developed in 1978 and has been adapted for use in other species, such as primates, dogs, cats, and horses (Ekman 2009; Wathan et al. 2015; Waller et al. 2020). It is a comprehensive catalogue of all potential facial configurations and is used as a research tool, allowing detailed measurement of facial expression in a standardised way.

Horses are a highly social species forming complex, long-term social relationships for which effective communication would be advantageous (Wathan et al. 2015). They have complex facial muscles, allowing for a wide range of facial movements or expressions, and rely heavily on visual perception (Wathan et al. 2015). The Equine Facial Action Coding System (EquiFACS) developed by Wathan et al. (2015) was developed through detailed anatomical examination and analysis of naturally occurring behaviour in coherence with other species' FACS. Most facial movements in horses are bilaterally symmetrical, except for ear and nostril lift movements. An ethogram has also been developed for facial expressions of ridden horses, which has been applied in combination with a pain scale to identify lame horses (Lezama-García et al. 2019).

The similarity of FACS between species has provided evidence of homology and evolutionary relatedness (Waller et al. 2020). FACS may allow for cross species comparison of facial behaviour, particularly if methods of analyses are developed to capture the complexity of movements and patterns, rather than individual action units. However, inference of emotion based on homology of expression with humans alone cannot be assumed to be valid (Waller et al. 2020). For example, while dogs show different, specific facial movements in response to different emotional stimuli, their expressions are not always analogous to those of humans in comparable situations (Lezama-García et al. 2019).

Manual recording of facial movements or expressions is labour and time intensive, somewhat subjective, and prone to human error. Tools such as FACS, kinematics, and computer image analysis involving computer learning (artificial intelligence), may contribute to the future ability to standardise and automate the process of accurately and objectively quantifying subtle changes in facial behaviour (Guesgen et al. 2017).

An attempt has been made to recognise emotions in horses using computer learning (convolutional neural networks) (Corujo et al. 2021). Emotions were defined by the authors as alarmed, curious, relaxed, or annoyed and were ascribed to a collection of still images of horses via subjective interpretation. No verification of emotion or validation of the ethogram was employed. The images were used to inform the computer model. Although the study does little to enhance the identification or measurement of emotion in horses, it does show that computer automation and interpretation is feasible in horses. Objective identification of emotional valence and arousal displayed in the images, rather than discrete emotions, is required to inform the model. The validity of interpretation of a still image captured in the field requires exploration. It is suggested that analysis of video may lead to more accurate assessment of emotion.

A study comparing the facial expressions of horses subject to either a dynamically adaptive, individualised, grooming treatment, inducing putatively positive emotion, to a standardised invariant one, inducing putatively negative emotion were found to differ with emotional valence, and found to be more sensitive than other behaviours when horses were re-exposed to grooming a year later (Lansade et al. 2018). The authors suggested that both groups were similarly aroused based on heart rate and concluded that the behavioural variation was a result of the opposing emotional valence. The study used a quasi-subjective selection of up to 20 still frames taken from video recordings of horses during treatment, time sampled at approximately 30 second intervals, depending on image quality. After removing data from facial behaviours that were not often observed (ears pinned back, low neck), a treatment effect was found in all measured facial behaviours (neck high/medium, eye white, eyes wide open/half-closed, lips straight/contracted/extended/mobile, ears asymmetrical), except for ears forwards and ears backwards. However, the study did not have an intermediate category, so whilst differences could be seen between the two treatments in reference to each other, causality could not be attributed to either treatment/emotional valence.

Principal component analysis (PCA) of the data revealed a difference between treatments for one factor, which represented 42% of the total variability (Lansade et al. 2018). It was negatively correlated with neck high, eye white, eyes wide open, lips contracted (better described as retracted lips), and ears asymmetrical, which were typical of horses in the fixed grooming/negative emotion group. The factor was positively correlated with neck medium, eyes half-closed, lips extended, lips mobile, and ears backward, which were typical of horses in the responsive grooming/positive emotion group. However, the method of behaviour measurement may have reduced the data variability and impacted the analysis. A year later the facial behaviours lips contracted, lips extended, and eye white continued to differ between groups in response to a short, fixed, but gentle brushing. The authors note that the facial expressions seen may be specific to the grooming context and the particular arousal levels, and that facial expression may be more sensitive than body behaviour for assessing emotion.

A scale for scoring sedation in horses was developed based on facial behaviour (De Oliveira et al. 2021). The degree of relaxation of the ears, the size of orbital opening, upper and lower lip relaxation were included in the scale. The implications of the study with respect to emotion are unclear.

During transportation of horses, a presumed negative stimulus, the rate of nostril dilation and upper eyelid raised was increased (Lundblad 2018). The author suggests that stress displayed in facial expressions may confound the use of facial expression when evaluating pain, although further research is required to provide support.

Eye

Eye aperture and visible sclera (eye white) may be an indicator of emotion. An increase in the proportion of visible eye white (sclera) may be an indicator in cattle of negative emotion (frustration), and a decrease is associated with low arousal positive emotion (contentment) (Sandem et al. 2002; Sandem et al. 2006; Proctor et al. 2015). It is presumed that negative emotion leads to sympathetic nervous stimulation induced retraction of the eyelids, which exposes more of the sclera. Similarly, in sheep, relative eye aperture was shown to be lowest with the positive stimulus and highest with the negative one, with the intermediate stimulus falling in between (Reefmann et al. 2009a). However, contrasting results have also been found in cattle with low eye white associated with removal of a negative stimulus, and increased eye white when eating a preferred feed (Whittaker et al. 2019). It may be that eye white, or eye aperture, is an indicator of arousal rather than valence.

Eyebrow wrinkles may provide information on emotional valence. Grooming (putatively pleasant) was associated with a flattening of the eyebrow wrinkles relative to the eye in horses (Hintze et al. 2016), although other measures were not significantly changed. However, the number and angle of eyebrow wrinkles was not found to differentiate emotional state in a judgment bias test in horses (Hintze et al. 2021). Further research is required to investigate the relationship of eyebrow wrinkles and eyebrow angle and emotion (Whittaker et al. 2019).

The rate of eye blinks decreased in horses with stress induced by feed restriction, a startle test, and social separation, compared to control (Merkies et al. 2019). Additionally, the rate of eyelid twitches increased in the feed restriction treatment. However, the results may indicate arousal rather than valence.

In humans, the startle response, a universal reflex characterised by a blink and tensing of the neck and back muscles, is thought to increase in amplitude with unpleasant stimuli and decrease with pleasant stimuli, relative to neutral stimuli (Mauss et al. 2009). It may represent emotional valence, decreasing with more positive valence, but only in the context of high arousal.

Ear

Ear movement frequency and position have been investigated as indicators of positive emotion in several species. An increase in ear movement frequency was associated with negative emotion from a less enriched environment in pigs (Marcet-Rius et al. 2019). However, in cattle, increased ear movement was associated with stroking, a putatively low arousal positive emotion (Whittaker et al. 2019).

In sheep, passive ear position (ears hanging down loosely) was associated with putatively low arousal positive emotion, whereas ears forward were seen with putatively negative emotion (Reefmann et al. 2009b; Reefmann et al. 2009a). The proportion of axial ear positions (ear out to side) was highest with a positive stimulus (voluntary grooming by a familiar human) when compared to a negative stimulus (social isolation). An intermediately valenced stimulus (standing in feeding area) was associated with an intermediate level of axial ear positions. The proportion of forward ear movements also tended to decrease from negative, through intermediate, to positive. The rate of all ear movements was highest with the negative stimulus (Reefmann et al. 2009a). Another study found that sheep in a presumed neutral state have their ears horizontal (axial), those in an unpleasant, uncontrollable situation have their ears pointed backwards, and those in an unpleasant but controllable one have their ears pointed upwards (Whittaker et al. 2019).

Ears flattened backwards have been associated with negative situations in dogs, pigs, and horses (Whittaker et al. 2019). Ears forwards is anecdotally associated with positive interest in horses, but has been found with negative vigilance in mice and dogs (Whittaker et al. 2019). Horses moved their ears forward in response to positively associated human vocalisations, and backward in response to negatively associated ones (d'Ingeo et al. 2019). However, in another study, horses' ear movements were not shown to vary with emotional arousal or valence, although it may have been confounded by directional attention to the stimulus, another horse, or horses leaving or returning (Briefer et al. 2015b).

Mouth

In horses reunited with a conspecific following a period of separation, the duration of non-masticatory chewing was found to decrease with increasing arousal level (as determined by heart rate), but was also affected by valence, with chewing increasing from negative to positive valence (Briefer et al. 2015b). The authors suggest that the non-masticatory chewing may relate to juvenile snapping or jaw champing behaviour, with possible self-calming or appeasement functions. Alternatively, it may also relate to the effect of the stress response on saliva release. It may be that a dry mouth is experienced with stress (e.g., from social separation), and an increase in lip licking or chewing movements seen post-stress (e.g., with social reunion) are to aid re-wetting of the lips and mouth, rather than being associated with an absolute positive valence per se (Thorbergson et al. 2016).

Dogs, horses, and cats have a play face that involves lip corner retraction with an open mouth, relaxed or stretched jaw, and teeth partially or totally covered, with some similarities to smiling in humans and primates (Descovich et al. 2017). However, there can be overlap with some signals of aggression or appeasement.

Changes in facial expression may be subtle and difficult to detect. In rats emitting 50kHz ultrasonic chirping when exposed to ‘tickling’ by humans, very little change in facial expression was found (Finlayson et al. 2016). The ears were less erect (wider angle, more relaxed, forward, and outward), and had darker pink colour than during the intermittent white noise exposure. The colour change may have been associated with arousal rather than valence. However, due to movement blur during treatment, only images that were taken after treatment were analysed. It is possible that assessment during treatment would reveal further information.

Many factors can influence assessment of facial behaviour from images. Lighting and shadows may affect appearance and colour assessment. Selection of images for analysis from video recordings or high frame rate still photographs may introduce bias. It may be better to analyse video rather than still images. A high-resolution video camera with a high frame rate may help provide detail and counter motion blur. It may be helpful to mark specific points on the face, or use a head mounted camera to stabilise the face image (Guesgen et al. 2017). Post-processing such as balancing and cropping of images may also have unintended impacts.

Care needs to be taken to validate the facial expression with other potential indicators of emotion, and to assess the accuracy of observers in coding or interpretation. As research in rabbits has shown, facial expression may not always reflect pain, and experienced and inexperienced observers may not be accurate at scoring pain from interpretation of facial expressions (Leach et al. 2011). Close range, high-resolution, high-quality video, or fast frame rate still images are necessary. Filming conditions, lighting, shadows, appearance of the animal, and presence and appearance of humans, may impact analysis of images. Coding video at reduced playback speed may be useful. Individual variation in facial morphology may have an impact. It is suggested that a neutral expression 'baseline' should be obtained prior to coding images (Descovich et al. 2017). Continued research on facial displays will yield valuable information on emotion in animals, as it has done in humans (de Waal 2011).

Approach-avoidance and behaviour tests

At a basic level, approach and avoidance behaviour may be used to indicate the desirability of a stimulus to an animal (Paul et al. 2005). Frequency and duration of approach (reward-seeking) and avoidance behaviours were used to verify the presumptive emotional valence in horses subjected to one of two grooming methods (Lansade et al. 2018). However, approach behaviour is also seen in negative contexts, such as defence of resources, and where escape is limited, in response to threat from predators.

Approach or avoidance behaviour has been assessed using behaviour tests such as the novel object test, voluntary human approach (non-interacting, passive human), and forced human approach tests, and have been utilised in studies on emotion with varying methods and results (Smith et al. 2016; Carreras et al. 2017; Riemer et al. 2018; Tamioso et al. 2018; Krugmann et al. 2019; Schrimpf et al. 2020). Standardised tests such as the open field, elevated plus maze, hole board, black and white box, forced swim test, startle, sucrose consumption, conditioned place preference, and social interaction tests have been used in mainly negative contexts (Paul et al. 2005). For a critical review of fear tests see Forkman et al. (2007). The startle response is enhanced when in a negative emotional state, but attenuated when in a positive one, and infers both arousal and valence (Paul et al. 2005).

Play and exploratory behaviour

Play behaviour has been investigated in several species. It may include functional elements, such as fighting, fleeing, and predatory behaviour, with much inter- and intra-species variation (Whittaker et al. 2019). Play behaviour often occurs once other needs, such as safety and hunger, have been met, and is commonly thought of as evidence of positive emotion and positive welfare (Paul et al. 2005; Hausberger et al. 2012; Keeling 2019). Spontaneous play behaviour may be scalable, at least in pigs, and the level of it used to differentiate between levels of good welfare (Keeling 2019). Controversially, it is suggested that current knowledge does not support the use of play to differentiate between good welfare and neutral welfare (Ahloy-Dallaire et al. 2018).

Play may also be a coping mechanism or response to restriction of resources, such as space, food, and social opportunities, or to situations of social conflict (Hausberger et al. 2012; Lesimple 2020). A study on play behaviour and measures of chronic stress in adult horses by Hausberger et al. (2012) contrasts with the commonly held view that the expression of play behaviour is an indicator of good welfare. However, further work is needed to support the authors' conclusion that play in adult horses is linked to poor welfare. An alternative explanation is that the high level of stereotypic behaviour seen may suggest that the horses' environment was poor regardless of play behaviour. Or it may be that in horses with stereotypies, play behaviour has a different function. Perhaps horses that are more motivated for play behaviour are more affected by space and social restrictions that limit their ability to perform it, thus the results are a consequence of thwarted motivation in some animals.

Play and exploratory behaviour are not very specific and are influenced by many factors, requiring careful interpretation. The definition and variability of play behaviours, and decline in behaviour with increasing age, complicate the use of play behaviour as an indicator of emotion (Ahloy-Dallaire et al. 2018).

Other behavioural parameters

Performance of specific behaviours, such as self-grooming, have been suggested to be indicative of pleasure (Keeling 2019), although may also be performed in negative contexts, such as skin disease or parasite infestation, or post-inhibitory rebound, and may be performed as a displacement behaviour when anxious, or as a prelude to agonistic encounters (Whittaker et al. 2019). Thus, careful consideration of the context and interpretation is required.

Behavioural diversity relates to the variety of behaviours and frequency of each behaviour that an animal displays (Miller et al. 2020). It is theorised that behavioural diversity may be useful as an indicator of positive emotional and welfare state (Miller et al. 2020). It is proposed that animals displaying a wide range of species typical behaviours have good welfare. This may be due to provision of resources and management that allows opportunities to engage in these behaviours, that is, environmental enrichment.

Qualitative behavioural assessment

Qualitative behavioural assessment (QBA) involves the subjective interpretation of photographic or video recordings of animals' behaviour by groups of people and assessing the degree of agreement of the descriptive terms used within clusters. It has shown moderate association with objective measures. However, criticisms include poor reliability (inter- and intra-observer), confounding from contextual cues, and its subjective nature (Webb et al. 2018). It is unclear if momentary emotions, or emotional states, can be assessed using QBA. In some instances observers have been shown to focus on body areas, which may not give an accurate measure of the emotion being studied, e.g. pain scoring in rabbits (Leach et al. 2011). Also referred to as free choice profiling, the method may be useful for behaviour and welfare assessment in species such as horses, dogs, cattle, pigs, and sheep (Wemelsfelder et al. 2001; Wemelsfelder 2007; Napolitano et al. 2008; Walker et al. 2010; Hintze et al. 2017).

3.5.3 Cognitive measures

Cognition is involved in both emotion and emotional state (Kremer et al. 2020). Cognition affects emotion and emotional state affects cognitive processing, thus they are interdependent (Paul et al. 2005). Cognitive measures of emotion may include cerebral and motor lateralisation, reward sensitivity, emotional balance, and forms of cognitive appraisal or bias such as judgement, memory, attention, and affective bias.

In a review, Leliveld et al. (2013) concluded that there is a similar cerebral lateralisation pattern for emotional processing across most vertebrate species. In support of the emotional valence hypothesis of emotional processing, while both cerebral hemispheres are involved, the right hemisphere dominates when processing negative emotions associated with fear or aggression, and the left dominates when processing positive emotion associated with a food reward. They suggest that hemispheric dominance, as evidenced by sensory and motor lateralisation of eyes, ears, nares, tail wag, and movement or limb preference behaviour, EEG, tympanic membrane temperature, or FNIRS, may be used as an indicator of emotional valence. However, supportive and contradictory results have been found, particularly with positive emotion (Quaranta et al. 2007; Farmer et al. 2010; McDowell et al. 2016; Smith et al. 2016; Wells et al. 2017; Siniscalchi et al. 2018; d'Ingeo et al. 2019; Giljov et al. 2019). It may be that cerebral lateralisation in animals reflects underlying approach-avoidance motivations, as it does in humans, with left hemisphere activation linked to approach and right hemisphere linked to avoidance states (Mauss et al. 2009). Alternative explanations for apparent cerebral lateralisation include the familiarity of stimuli and rapidity of response (Leliveld et al. 2013; Whittaker et al. 2019).

Cognitive appraisal may be influenced by emotion in three main areas: attention (increased vigilance with anxiety), memory (improved memory formation and recall with positive and negative emotions compared to neutral), and judgement (interpretation of ambiguous stimuli). Methodologies involving judgement bias, sensitivity to reward loss, and risk taking have been used to investigate emotional states and welfare status in animals with varying results (Kremer et al. 2020). Paul et al. (2005) provide a review of cognitive bias and its application for measuring emotion in animals. The methods may be time-consuming, and confounded by the experimental process, which may be a cognitively enriching experience in itself (Webb et al. 2018). Equivocal and unexpected results have been seen in some studies (Burani et al. 2020; Hintze et al. 2021). Although most often used in research on emotional states, it may be that cognitive bias is influenced by both momentary emotion as well as overlying emotional state, and potentially underlying individual traits such as temperament (Webb et al. 2018; Kremer et al. 2020).

Emotional balance

It has been suggested that emotional balance, calculated as the ratio of the frequency of positive to negative emotions, can be used to determine an absolute rather than a relative emotional state in animals (Boissy et al. 2007b; Webb et al. 2018; Kremer et al. 2020). In humans, it is the frequency with which positive and negative emotion is experienced, rather than the intensity, that is important to the affective happiness component of life satisfaction (Fredrickson 2013; Kremer et al. 2020). The concept of an emotional balance measure in animals requires validation with further research on detection and measurement of the momentary emotions needed to underpin it (Kremer et al. 2020).

3.6 General methodological considerations

There are many methodological issues to consider when researching indicators of emotion in animals (Kremer et al. 2020). Subject related factors may include sex, breed, age, reproductive status, genetics, previous experience, learning and memory, laterality or side bias, and other individual traits. Animals displaying abnormal repetitive behaviour may respond differently and should be considered separately from normal animals when assessing emotion (Lebelt et al. 1998; Normando et al. 2007).

Arousal level can confound assessment of valence. There may be an interaction between emotional arousal and valence that can be difficult to separate. Ideally, comparison should be made between emotions of a positive, neutral, and negative valence, while controlling for arousal level. It is challenging to design experiments involving a sufficient degree of opposing valence with the same level of arousal (Whittaker et al. 2019). There can be difficulty in distinguishing low arousal positive emotion from neutral and/or low arousal negative emotion (Kremer et al. 2020). Additionally, while a neutral valence is a theoretical possibility, in reality it may be a point that is indeterminate, individually unique, and relative, rather than absolute. Perhaps use of the term intermediate valence, being in-between the most positive and most negative emotion, is a more realistic term.

Much progress has been made on identifying negative emotions, such as pain, fear, and frustration, which are often associated with high arousal. Signals of positive emotion may be less obvious. Measures need to be sufficiently sensitive to be able to detect subtleties of low arousal positive emotion. Individual variation may overshadow subtle changes, requiring careful attention to study design, sample size, and analysis.

With induction of emotion in the research setting, anxiety or fear may overshadow the intended emotion without careful attention to factors such as familiarity, choice, and restraint or ability to escape. Familiarity and past experiences with the experimental environment, treatments, and cues, as well as the experimenters, may influence results. To avoid potential impact from novelty, subjects may need to be familiar with the positive emotion inducing stimuli prior to testing. Social species may be more affected by isolation or disruption to social groupings, so a conspecific(s) may need to be present. It may be unknown whether any conspecific will suffice, or if a particular bond, or agonistic relationship, exists between herd mates. Similarly, as a communicative function, emotional signals may require the presence of an appropriate audience to be displayed. The possible effects, intended or otherwise, of emotional transfer within and between species should also be considered.

The data collection process may influence results. This may be obvious with invasive sampling (e.g., blood collection), but there may also be effects from the presence of equipment, and extraneous sensations. The time course and temporal relationship of emotion inducing stimuli and measurements should be considered. Questions such as appropriate duration of stimulus, when to measure, lag time between the stimulus and response, how long to measure for, if/when to measure baseline values, and if/when to measure return to baseline values, are important, as are factors such as cyclical patterns and time of day effects on individual baseline values (Kremer et al. 2020). An appropriate baseline may need to be defined, but baseline values may drift over time, independent of treatment sessions, especially if habituation or sensitisation is occurring.

3.7 Conclusion

Limited research has been conducted on positive emotion inducing experiences and the assessment of positive emotion in horses. Some parameters, whilst giving valuable information, would be impractical, too invasive, or costly to investigate in horses. Some parameters require further investigation before field application in animals. Other parameters, such as assessment of cognitive bias, may be more suited to the study of emotional states. It seems likely that no single measure will be suitable for assessing emotion in animals, and the use of a single measure may lead to erroneous conclusions. The use of multiple measures, for example, the combination of behaviour analysis with physiological data, is likely to add confidence to the interpretation of results.

Challenges include induction of emotion, control of confounds, and disentangling emotional arousal and valence. Some physiological changes may reflect the intensity of arousal or physical activity, not necessarily emotional valence (Paul et al. 2005). The same behaviour performed in different contexts may imply a different underlying emotion or motivation. (Kremer et al. 2020). Interpretation of physiological, behavioural, or cognitive parameters in the context of emotion is complex and requires a cautious approach to avoid bias.

3.8 Preface to experimental chapters: Research approach

The information in this review helped to inform the thesis research approach. Having identified potentially useful methodologies and gaps in knowledge, an initial experiment was conducted, from which two main issues were identified. Horses were subjected to experiences that were assumed to be pleasant or unpleasant, and various physiological and behavioural measures were recorded to determine their potential as indicators of emotion (Chapter 4).

This approach was flawed. Firstly, the assumption that the experiences would be universally pleasant or unpleasant did not hold up. Some horses appeared to have contrary reactions to the putatively positive and putatively negative treatments. The emotional reaction of individual horses is a product of their genetics, environment, and their prior experience or learning, and is inherently unique to the individual. What one horse finds pleasant, another may not.

The second issue involves determination of the valence of the emotion induced by a stimulus. In the absence of an established gold standard method, many animal studies rely on pre-assumption of valence from perceptions of what animals do or don't like, or the similarity of behavioural and/or physiological responses to those in humans (Seaman et al. 2002; Briefer Freymond et al. 2014; Briefer et al. 2015b). There is risk of a fundamental circular argument, whereby the potential indicators of the putatively positive or negative emotions are also used to declare that the valence of the emotion is in fact positive or negative (Lawrence et al. 2019). A researcher may set out to measure positive emotion, but how can it be known when the animal is experiencing this emotion, and that they are measuring what they purport to measure? Studies on indicators of emotion should independently verify how the putatively positive or negative emotion inducing experience is actually perceived by the individual.

The subsequent research approach involved combining the use of preference testing to identify an individual's favoured stimulus, and examining potential indicators of emotion when experiencing this favoured stimulus, to resolve the arguments outlined above (Chapters 5-10). It was based on the premise that a preference for one stimulus over another is likely to mean that it is more pleasurable, and thus induces a more positive emotion (Humphrey 1972; Boissy et al. 2007b; Dawkins 2008).

A method of preference testing was designed for the second experiment to identify horses that demonstrated a consistent preference for one of the stimuli offered (Chapter 5). The third experiment, introduced in Chapter 6, investigated potential non-invasive physiological (Chapter 7), gross body behavioural (Chapter 8), and facial movement (Chapter 9) indicators of emotion, and how they varied with the preferences displayed in the second experiment. The results were then interpreted in terms of the arousal and valence dimensions of emotion (Chapter 10). Multivariate analysis was conducted to assimilate and simplify the indicators of emotion into underlying clusters (Chapter 11).

Chapter 4 Observation of physiological and behavioural responses to presumed positive and negative emotion inducing stimuli

4.1 Introduction

It is generally accepted that animals, at least mammals, experience emotions, both positive, such as pleasure and happiness, and negative, such as pain, fear, and frustration (Chapter 2). An emotion can be broadly defined as an innate, intense, but short-lived response to an event (internal or external) that has behavioural, physiological, subjective, and cognitive components (Panksepp 2005; Paul et al. 2005; Boissy et al. 2007b; Garland et al. 2010; de Waal 2011). Unlike humans that can verbally communicate their emotions, it is not possibly to definitively know what subjective feeling an animal is experiencing. However, it may be possible to infer it from behavioural and physiological changes (Boissy et al. 2007b).

Objective assessment of positive emotion may aid in the provision of circumstances to enhance welfare in animals. Many animal welfare standards emphasise the absence of negative emotions, such as pain and fear. A recent focus of animal welfare science is improved welfare or wellbeing, which includes pleasurable experiences, as well as the absence of unpleasant ones (Boissy et al. 2007a; Boissy et al. 2007b; Fraser 2008; Yeates et al. 2008; Balcombe 2009; Green et al. 2011; Mellor 2015; Webb et al. 2018; Keeling 2019; Lawrence et al. 2019). It is necessary to determine what an animal enjoys and wants, as opposed to what it doesn't like and would rather avoid. Preference and avoidance testing are means of achieving this (Appleby et al. 2011). The ability to identify situations perceived as positive by horses would allow us to improve their health and welfare.

In this experiment, horses were exposed to stimuli thought likely to induce positive and negative emotions, while selected physiological and behavioural variables were measured. The aim was to investigate potential non-invasive behavioural and physiological indicators of emotion in horses experiencing putatively pleasant and putatively aversive treatments. The assumption was made that a horse undergoing a pleasant treatment will be experiencing positive emotion, and conversely that an aversive treatment will elicit negative emotion.

4.2 Materials and methods

All procedures were conducted in accordance with the Massey University Code of Ethics for Animals (MUAEC approval 09/96). The experiment was conducted at the Massey University Veterinary Large Animal Teaching Unit, Manawatu, New Zealand. The facilities used included covered yards and indoor stocks. The experiment was conducted in early summer.

4.2.1 Experiment overview

Following a habituation and training phase, thirteen horses were subjected to each of four treatments: putatively pleasant grooming (wither massage), putatively aversive grooming (inguinal scrub), exposure to an auditory secondary reinforcer (clicker, also putatively pleasant), and control (where no additional stimulus was applied), in a cross over design. Each horse was exposed to each treatment once, with ~24 hours between each treatment to minimise carryover effects. Measures of heart rate (HR), heart rate variability (HRV), respiratory rate (RR), eye/ocular surface temperature (ET), skin temperature (ST), and relative humidity above the skin (RH), were continuously recorded prior to (10 minutes, Pre-10to-5, Pre-5to0), during (5 minutes, Treat0to5), and after (10 minutes, Post5to10, Post10to15) the treatments. Video was continuously recorded and was used for coding of movement behaviour. A timeline for each treatment session is displayed in Figure 3.

Attempts were made to control for potential confounding factors. Horses were habituated to the experimental facilities, equipment and personnel, and had prior exposure to the putatively pleasant grooming technique (McBride et al. 2004). They had a conspecific nearby. A modified Latin square was used to balance the day and order of treatment effects, the time of the day was kept consistent for each horse, treatments were applied by three familiar people (all female) that had not had any negative interactions with the horses, habituation and training was conducted using mainly positive reinforcement, and consistency in application of treatments was maintained. When initially led in, no animals displayed signs of fear or aversion to the experimenters or to the stocks (an open chute made from pipe stanchions, a single rail on each side, and a single rope in front and behind the horse to provide loose restraint, see Figure 18). Horses were backed into the stocks, loosely cross-tied, and the equipment was then attached to them. They were given time to acclimate before recording commenced, and during the first 5 minutes of recording (period Pre-10to-5). This acclimatisation period was not included in the analyses.

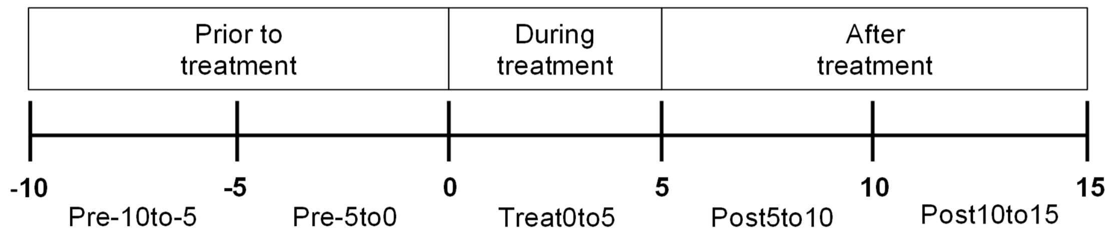


Figure 3 Timeline for each treatment session.

4.2.2 Animals and housing

Thirteen horses, six of which were kept by the university and seven of which were privately owned, were used in the experiment. Due to the novelty of the experiment, an *a priori* power analysis could not be performed. The number of subjects was based on availability, suitability, ethical, and practical constraints (Taborsky 2010).

Horses were kept at pasture in established groups so that social bonds were not disrupted. Horses that were not normally resident at the facility were transported by road trailer (approximately 20km) prior to the experiment. Horses were walked approximately 0.5km daily from paddocks to the testing facility.

Horses were accustomed to being handled. Their diet consisted of grass and small amounts of hay with water provided *ad libitum*. Horse ages ranged from 5-20 years old. There were 10 geldings, two mares, and one ovariectomised mare; five were Standardbred, six were Thoroughbred or Thoroughbred cross, and two were mixed breeds (Table 1).

Horses were excluded from the study if they displayed a fear response to the control stimulus, were affected by illness or injury during the study, or if they displayed stereotypic or abnormal behaviour.

Table 1 Subject details.

Group	ID	Name	Sex	Age	Breed	Treatment order
1	1	Justin	G	5	SB	C S M K
	2	Shane	G	5	SB	S C K M
	3	Dougal	G	20	TB	M K C S
	4	Fatty	G	10	SB	K M S C
	5	Zinga	M	7	TBx	C S M K
	6	Tiger	G	10	TB	M K C S
2	7	Churchill	G	8	TB	K M S C
	8	Tucker	G	*	SB	S C K M
	9	Boromir	G	6	Mixed	C S M K
	10	Gem	M	9	TB	K M S C
	11	Sovereign	G	9	Mixed	S C K M
	12	Grace	Ovx	16	TBx	M K C S
	13	Morrie	G	5	SB	C S M K

G = Gelding, M = Mare, Ovx = Ovariectomised Mare, SB = Standardbred, TB = Thoroughbred, x = Cross, * = unknown, C = Control, S = Scrub, M = Massage, K = Clicker

4.2.3 Treatments

The treatments used to elicit emotion are described in Table 2, and were selected based on review of experimental literature (Feh et al. 1993; Normando et al. 2003; McBride et al. 2004; Burgdorf et al. 2006; Normando et al. 2007; Reefmann et al. 2009a). The wither region is a preferred site for allogrooming, and tactile stimulation (e.g., massage, brushing, grooming) of it may have a pleasant relaxing effect in horses (Feh et al. 1993; McBride et al. 2004). To avoid the potential impact of social isolation, a familiar companion horse (not a test subject) was placed in the stocks in front.

Table 2 Name and description of treatments.

Treatment	Putative emotion	Description
Control (C)	Neutral	Horses stood in stocks, loosely cross-tied. No additional stimulus was present.
Scrub (S)	Negative	A person standing on the left side vigorously rubbed and tapped the ventral abdominal and inguinal areas with long handled stiff scrubbing brush.
Massage (M)	Positive	A different person rubbed the left withers and shoulder area using a soft rubber bristled grooming tool in a firm, slow, circular motion with consistent pressure.
Clicker (K)	Positive	A different person standing beside the shoulder activated a clicker sound (previously classically conditioned with food) every 2 seconds. Horses were not restricted from nuzzling/investigation of experimenter unless biting was attempted, in which case the person stepped backwards.

4.2.4 Measures

Measures were chosen based on the experimental literature on the assessment of positive emotions (Knutson et al. 2002; Stewart et al. 2005; Boissy et al. 2007b; von Borell et al. 2007; Stewart et al. 2008b; Stewart et al. 2008a; Reefmann et al. 2009c; Reefmann et al. 2009a; Stewart et al. 2010a). Measurements were recorded for a period of 10 minutes pre-treatment, during the 5-minute treatment period, and for 10 minutes post-treatment.

Heart rate and inter-beat interval (IBI) were continuously recorded using Polar Equine heart rate monitors (S810iTM, Polar Electro Oy, Helsinki, Finland), with the electrode strap placed around the chest and conductive gel on the sensors. Polar ProTrainer 5 Equine Edition (version 5.35.161, Polar Electro Oy, Helsinki, Finland) software was used to extract data. Error correction was performed using a moderate filter and a minimum protection zone of six beats per minute.

HRVanalysis (version 1.1, https://anslabtools.univ-st-etienne.fr/en/index.html) software was used to calculate measures of HRV from the IBI data including root mean square of successive differences (RMSSD), standard deviation of normal-to-normal beats (SDNN), and the ratio of RMSSD to SDNN.

Respiration rate was continuously recorded using a pressure transducer attached to a custom-made flexible tube around the thorax. Physiological data analysis software (Chart 5 version 5.5.6, ADInstruments Ltd, Dunedin, New Zealand) was used to extract the RR data.

Thermal images of the right eye were recorded continuously using an infrared thermography (IRT) video camera (ThermaCam S60, FLIR Systems AB, Danderyd, Sweden) positioned 1-2 metres from the right side of the horse’s head to measure ET. The maximum temperature at the medial canthus/lacrimal caruncle of the right eye was manually recorded. Segments of the thermography video were examined to provide verification of manual recording and additional data points as required, using image analysis software (ThermaCAM Researcher Pro version 2.10, FLIR Systems AB, Danderyd, Sweden). Clear images that contained a minimum of five pixels were used.

Skin temperature and RH above the skin were recorded once per second using a data logger (MSR145W, MSR Electronics GmbH, Switzerland), with the temperature sensor contacting shaved skin on the chest 10cm caudal to the right elbow, and the humidity sensor held 2-5mm above the shaved skin, covered by a breathable dressing.

Behaviour was recorded using four digital video cameras (Sony DCR-SR85E, Auckland, New Zealand). Cameras were positioned to provide views of the whole horse (front-right, rear, right-side), and a close-up of the head (front-left). Movements of the head, forelegs, and hind legs were coded according to the ethogram shown in Table 3. One observer scored all the videos. Video viewing software (Picture Motion Browser, version 5.2.00.03250, Sony Corporation, Tokyo, Japan) was used to trim videos to the required length. Coding was performed using behaviour data capture software (J-watcher With Video, version 1.0, Dan Blumstein, University of California Los Angeles & The Animal Behaviour Lab, Macquarie University, Sydney).

Table 3 Ethogram of movement behaviours recorded from video. Exclusions are behaviours which were incompatible with each behaviour, for example, a horse could not be said to lift its tail and swish it at the same time.

Body part	Behaviour	Exclusions	Description
Foreleg	Lift	All other foreleg	Complete lift of foreleg hoof from ground in a vertical direction
	Strike	All other foreleg	Complete lift of foreleg hoof from ground also with leg movement forward
	Point	All other foreleg	Heel bulb of foreleg removed from ground while toe of hoof remains in contact with ground
Hindleg	Lift	All other hindleg	Complete lift of hindleg hoof from ground in a vertical direction
	Kick	All other hindleg	Complete lift of hindleg hoof from ground surface also with leg movement backwards
	Rest	All other hindleg	Heel bulb of hindleg removed from ground while toe of hoof remains in contact with ground
Neck-head	Very low	All other neck positions	Neck lowered approximately -30 to -90º from horizontal through vertebral column
	Low	All other neck positions	Neck lowered approximately -10 to -30º from horizontal through vertebral column
	Neutral	All other neck positions	Neck is approximately in line with the withers and croup (horizontal with the vertebral column +/-10º)
	High medium	All other neck positions	Neck approximately 10-45º above the horizontal through vertebral column
	High	All other neck positions	Neck approximately 45-90º above horizontal through vertebral column
Tail	Lift	All other tail	Movement of dock in vertical plane resulting in dock placement higher than horizontal to spine/pelvis - including tail lift before elimination
	Swish	All other tail	Lateral movement of the dock without moving in a vertical plane away from the hindquarters
	Thrash	All other tail	Strong repeated vertical dock movement from the hindquarters with a gap clearly visible between the dock and hindquarters, with or without lateral movement
Elimination	Urination	All leg positions	Horse urinates, leg position not counted during urination
	Defecation	All leg positions	Horse defecates, leg position not counted during defecation

4.2.5 Habituation and training

Horses were trained over three sessions using food (as a lure and as positive reinforcement) and pressure (on halter and legs used as negative reinforcement) to calmly back into the stocks (to facilitate camera angles) and stand loosely cross-tied while wearing a surcingle (simulating the respiratory and heart rate monitor straps). Systematic desensitisation and counterconditioning, using food, were used to habituate the horses to the facilities, proximity of cameras, and to being shaved (for placement of the temperature and humidity sensors). Horses were also introduced to the wither massage treatment during the training phase to provide prior exposure and avoid fear from novelty. They were trained via classical conditioning to associate a clicker noise with the presentation and ingestion of a palatable food (sliced carrot or pellets depending on horse preference) (McGreevy et al. 2018).

4.2.6 Data Handling and Analysis

Continuous variables (HR, RR, ET, ST, RH, HRV measures: IBI, SDNN, RMSSD, and RMSSD to SDNN ratio) were summarised by mean and standard error (SE), stratified by treatment and period. Outliers were identified from assessment of boxplots, checked for errors, and corrected or kept as valid data. Normality was assessed at each level by Shapiro-Wilk test ($p > 0.05$). A series of one-way repeated measures ANOVA (Analysis of Variance) tests were conducted to determine whether there were statistically significant differences in each variable due to (i) treatment within each period, and (ii) period within treatment. However, the focus was how each treatment changed over time when compared to pre-treatment. The alternate analysis, comparison of treatments within a period does not account for pre-treatment/baseline levels.

Overall, analysis was conducted on four treatments and three to four periods, depending on outcome variable. Sphericity was assessed using Mauchly's test and, where the assumption was violated, the Greenhouse-Geisser correction was used (Laerd Statistics 2015).

The number of occurrences of individual behaviours was aggregated by body region (head/neck, foreleg, hindleg, tail, urination/defecation), stratified by period and treatment. The distribution of the data was assessed graphically as histograms and boxplots, and was non-parametric. The Friedman's two-way analysis of variance test was used to assess main effects of treatment or period. Overall, analysis was conducted on four treatments and four periods. For behaviour data, values given are median and range unless stated.

Statistical analyses were performed in IBM SPSS Statistics for Windows (Version 27.0. IBM Corp. Armonk, New York, USA). The statistical significance level was set at $p \leq 0.05$.

Bonferroni correction of p-values was applied during post-hoc pairwise testing to account for repeated measures.

4.3 Results

4.3.1 Heart rate

The results of analysis of period within each treatment are presented in Figure 4. Within each treatment, HR did not vary significantly by period (Control p = 0.265, Scrub p = 0.423, Massage p = 0.175, Clicker p = 0.315). Within each period, HR did not vary significantly by treatment (Pre-5to0 p = 0.967, Treat0to5 p = 0.061, Post5to10 p = 0.293, Post10to15 p = 0.430).

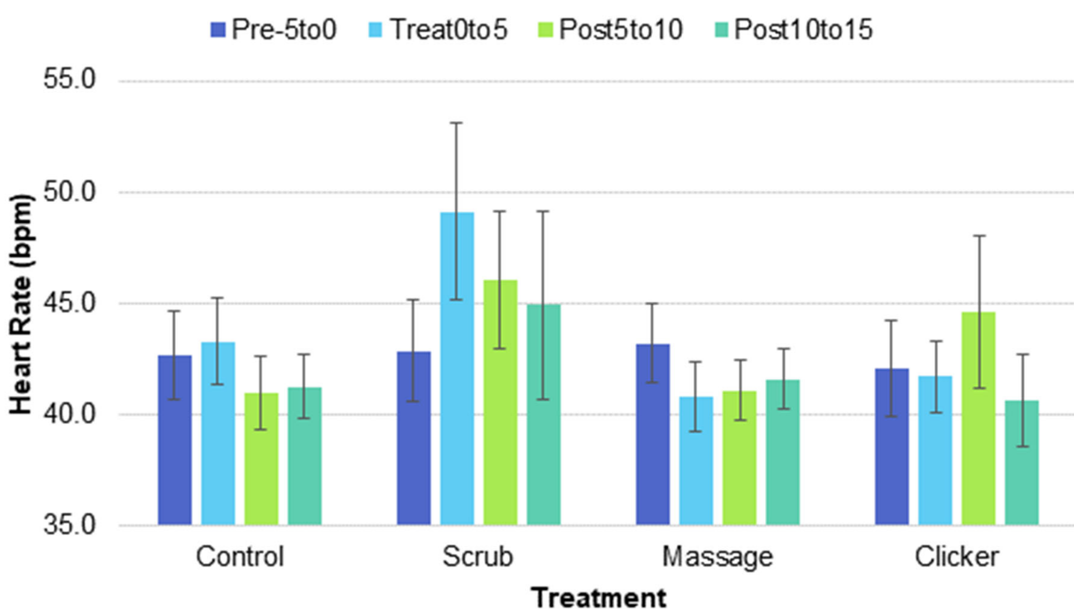


Figure 4 Estimated mean heart rate by treatment, clustered by period.
Error bars show SE. No significant differences between means were found within each treatment.
bpm = beats per minute. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

4.3.2 Heart rate variability

There was insufficient IBI data for one horse for the Post5to10 period of the Control and Massage treatments.

Inter-beat interval

Within each treatment, IBI did not vary significantly by period (Control p = 0.266, Scrub p = 0.135, Massage p = 0.089, Clicker p = 0.952).

For IBI during the Treat0to5 period, there was a significant overall effect of treatment (p = 0.028). No significant differences between treatments were found during post-hoc pairwise testing. For IBI during the other periods, there was no significant overall effect of treatment (Pre-5to0 p = 0.986, Post5to10 p = 0.862).

SDNN

For SDNN during the Scrub treatment, there was a significant overall effect of period (p = 0.012). No significant differences between treatments were found during post-hoc pairwise testing. For SDNN during the other treatments, there was no significant overall effect of period (Control p = 0.825, Massage p = 0.309, Clicker p = 0.774).

Within each period, SDNN did not vary significantly by treatment (Pre-5to0 p = 0.707, Treat0to5 p = 0.604, Post5to10 p = 0.157).

RMSSD

Within each treatment, RMSSD did not vary significantly by period (Control p = 0.908, Scrub p = 0.319, Massage p = 0.315, Clicker p = 0.765). Within each period, RMSSD did not vary significantly by treatment (Pre-5to0 p = 0.773, Treat0to5 p = 0.487, Post5to10 p = 0.175).

Ratio RMSSD/SDNN

For RMSSD/SDNN ratio during the Scrub treatment, there was a significant overall effect of period (p = 0.022). A significant difference between the Treat0to5 and Post5to10 periods (p = 0.033), with a mean of 0.791 (SE 0.079) and 0.957 (SE 0.032) respectively, was found during post-hoc pairwise testing. For RMSSD/SDNN ratio during the other treatments, there was no significant overall effect of period (Control p = 0.846, Massage p = 0.369, Clicker p = 0.928).

Within each period, RMSSD/SDNN ratio did not vary significantly by treatment (Pre-5to0 p = 0.954, Treat0to5 p = 0.104, Post5to10 p = 0.370).

4.3.3 Respiratory rate

There were no RR data for one horse for the Treat0to5 period of the Scrub treatment due to technical issues.

The results of analysis of period within each treatment are presented in Figure 5. For RR during the Scrub treatment, there was a significant overall effect of period (p = 0.003). No significant differences between treatments were found during post-hoc pairwise testing.

For RR during the Clicker treatment, there was a significant overall effect of period (p = 0.009). A significant difference between the Treat0to5 and Post5to10 periods (p = 0.004), with a mean of 13.54 breaths per minute (bpm) (SE 1.03) and 12.83 bpm (SE 1.22) respectively, was found during post-hoc pairwise testing.

For RR during the other treatments, there was no significant overall effect of period (Control p = 0.832, Massage p = 0.326).

For RR during the Treat0to5 period, there was a significant overall effect of treatment (p = 0.044). No significant differences between treatments were found during post-hoc pairwise testing. For RR during the other periods, there was no significant overall effect of treatment (Pre-5to0 p = 0.974, Post5to10 p = 0.667, Post10to15 p = 0.503).

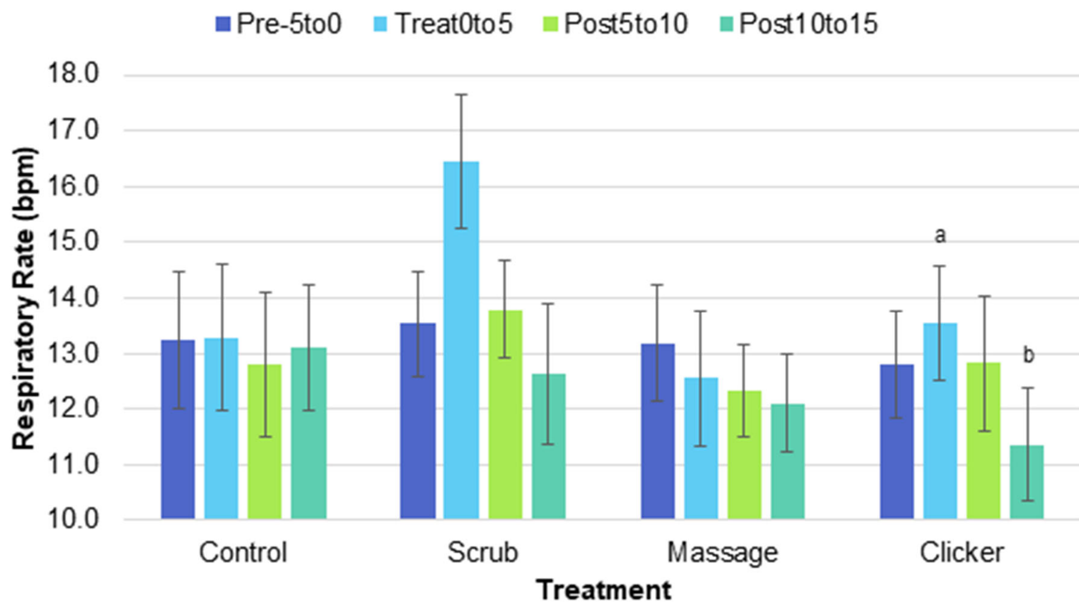


Figure 5 Estimated mean respiratory rate (RR) by treatment, clustered by period. Error bars show SE. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Clicker treatment mean RR for the Treat0to5 period was significantly different to the Post10to15 period (p = 0.009).
bpm = breaths per minute. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

4.3.4 Eye temperature

The results of analysis of period within each treatment are presented in Figure 6. For ET, during the Control treatment there was significant overall effect of period (p < 0.001). During post-hoc pairwise testing, significant differences were found between the Pre-5to0 and Post5to10 periods (p = 0.012), with a mean of 35.77°C (SE 0.12) and 35.98°C (SE 0.15) respectively, and the Pre-5to0 and Post10to15 periods (p = 0.005), with a mean of 35.77 °C (SE 0.12) and 36.08 °C (SE 0.15) respectively.

For ET during the Massage treatment, there was significant overall effect of period ($p = 0.005$). During post-hoc pairwise testing, significant differences were found between the Pre-5to0 and Post5to10 periods ($p = 0.026$), with a mean of 35.15^{o}C (SE 0.11) and 35.75^{o}C (SE 0.11) respectively. The mean ET during the Post10to15 period was 35.75^{o}C (SE 0.11), which differed to the Pre-5to0 ($p = 0.053$).

For ET during the Clicker treatment, there was a significant overall effect of period ($p = 0.017$). No significant differences between treatments were found during post-hoc pairwise testing. There was no significant effect of period on ET with the Scrub treatment ($p = 0.382$).

Within each period, ET did not vary significantly by treatment (Pre-5to0 $p = 0.352$, Treat0to5 $p = 0.713$, Post5to10 $p = 0.356$, Post10to15 $p = 0.334$).

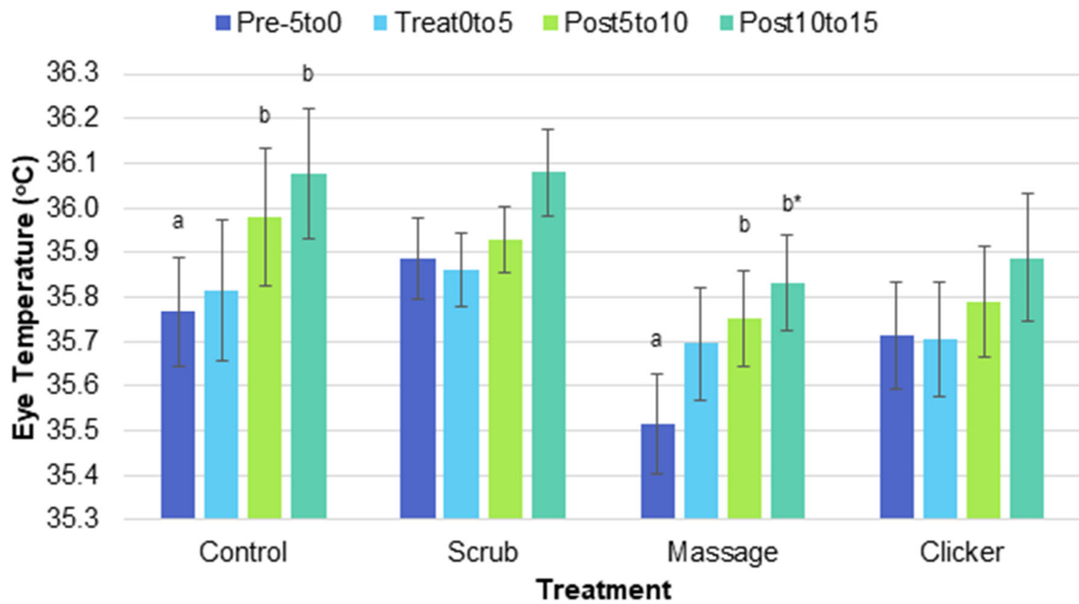


Figure 6 Estimated mean eye temperature (ET) by treatment, clustered by period. Error bars show SE. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Control treatment, mean ET for the Pre-5to0 period was significantly different to the Post5to10 (p = 0.012) and Post10to15 (p = 0.005) periods. Within the Massage treatment, mean ET for the Pre-5to0 period was significantly different to the Post5to10 (p = 0.026) and Post10to15 (p = 0.053) periods.
°C = degrees Celsius. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10. * p = 0.053

4.3.5 Skin temperature

Skin temperature data was missing for three horses for the Control and Scrub treatments, and another three horses for the Massage and Clicker treatments.

The results of analysis of period within each treatment are presented in Figure 7. For ST during the Control treatment, there was a significant overall effect of period (p = 0.005).

For ST during the Massage treatment, there was a significant overall effect of period (p = 0.024). During post-hoc pairwise testing, significant differences were found between the Post5to10 and the Post10to15 periods (p = 0.031), with a mean of 34.46°C (SE 0.26) and 34.64°C (SE 0.28) respectively.

For ST during the other treatments, there was no significant overall effect of period (Scrub p = 0.287, Clicker p = 0.103).

Within each period, ST did not vary significantly by treatment (Pre-5to0 p = 0.422, Treat0to5 p = 0.828, Post5to10 p = 0.303, Post10to15 p = 0.285).

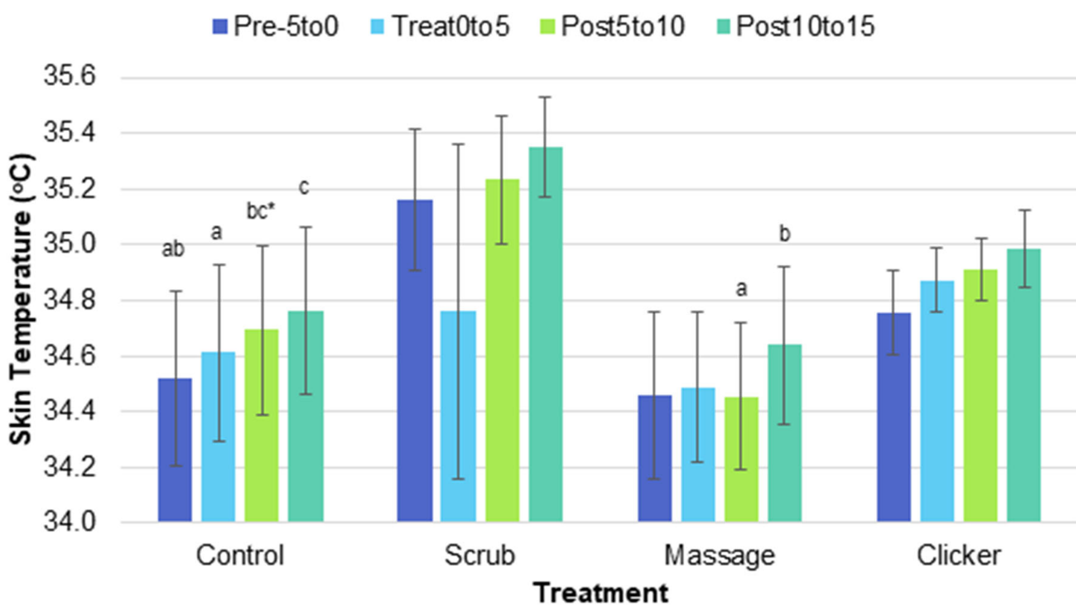


Figure 7 Estimated mean skin temperature (ST) by treatment, clustered by period. Error bars show SE. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Control treatment, mean ST for the Pre-5to0 period was significantly different to the Post10to15 period (p = 0.037), the Treat0to5 period was significantly different to the Post5to10 period (p = 0.053) and the Post10to15 period (p = 0.024). Within the Massage treatment, mean ST for the Post5to10 period was significantly different to the Post10to15 period (p = 0.031).
ºC = degrees Celsius. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10. * p = 0.053

4.3.6 Relative humidity above skin

Data was missing for three horses for the Control and Scrub treatments, and another three horses for the Massage and Clicker treatments.

The results of analysis of period within each treatment are presented in Figure 8. Within each treatment, RH did not vary significantly by period (Control p = 0.853, Scrub p = 0.522, Massage p = 0.330, Clicker p = 0.701). Within each period, RH did not vary significantly by treatment (Pre-5to0 p = 0.599, Treat0to5 p = 0.071, Post5to10 p = 0.118, Post10to15 p = 0.445).

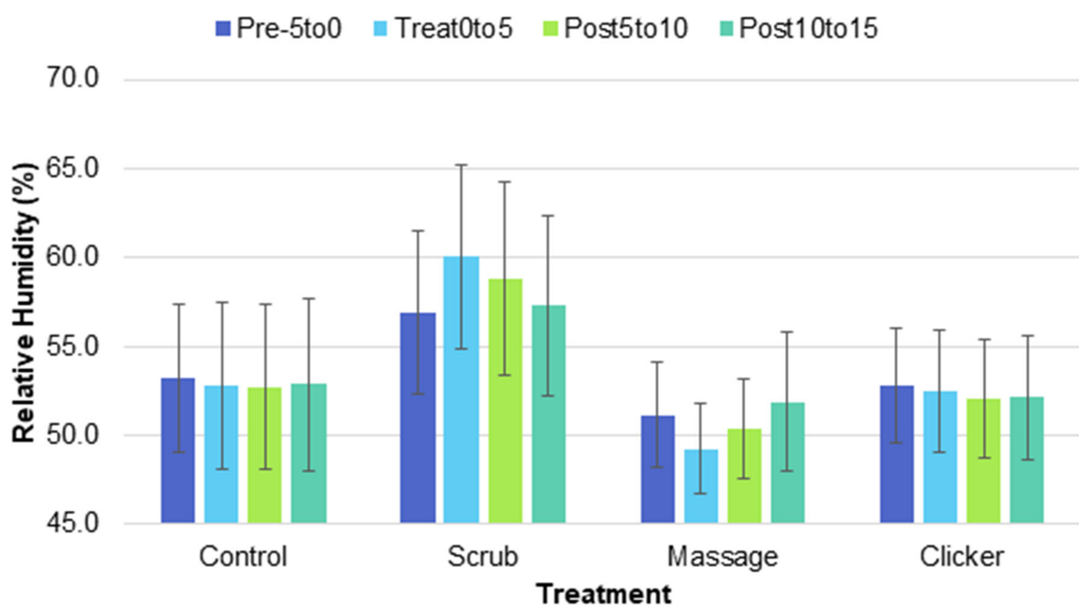


Figure 8 Estimated mean relative humidity above the skin by treatment, clustered by period.
Error bars show SE. No significant differences between means were found within each treatment.
% = per cent. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

4.3.7 Behaviour

Neck-head movements

The number of neck-head movements per period for each treatment is presented in Figure 9. For neck-head movements during the Massage treatment, there was a significant overall effect of period (p = 0.048). A significant difference between the Treat0to5 and Post10to15 periods (p = 0.037), with a median and range of 20 (1-85) and 57 (18-147) neck-head movements respectively, was found during post-hoc pairwise testing.

For neck-head movements during the Clicker treatment, there was a significant overall effect of period (p = 0.039). No significant differences between treatments were found during post-hoc pairwise testing. For neck-head movements during the other treatments, there was no significant overall effect of period (Control p = 0.854, Scrub p = 0.067).

Within each period, the number of neck-head movements did not vary significantly by treatment (Pre-5to0 p = 0.889, Treat0to5 p = 0.313, Post5to10 p = 0.721, Post10to15 p = 0.132).

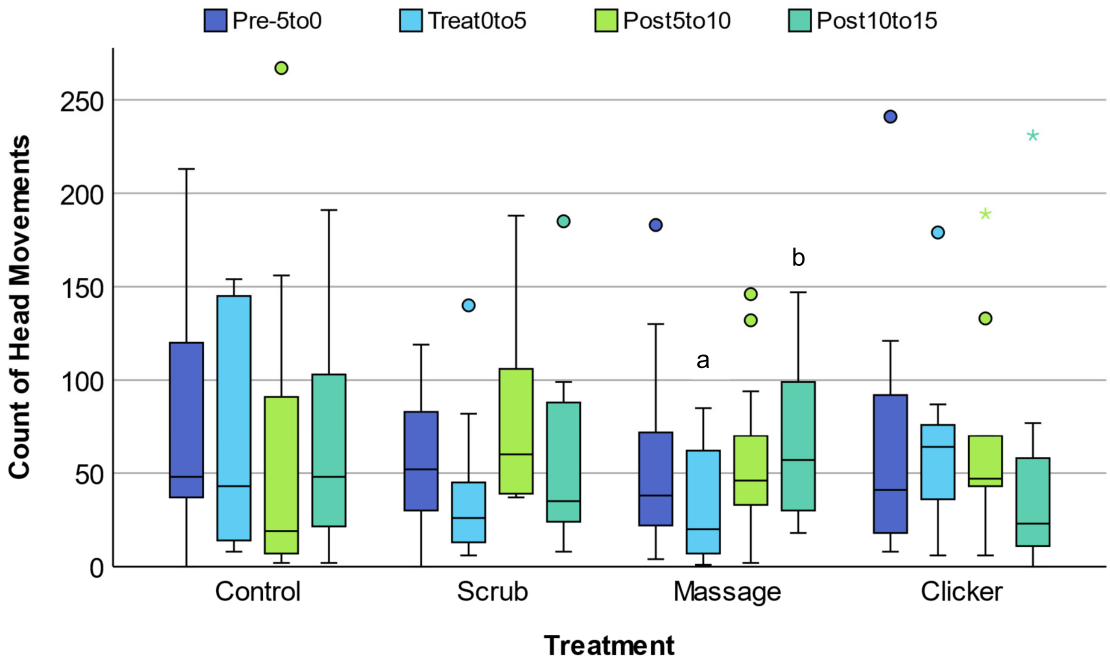


Figure 9 Boxplot of neck-head movements by treatment, clustered by period. Circles indicate potential outliers (1.5-3*IQR), and asterisks indicate extreme outliers (>3*IQR). Some outliers have been removed to aid visualisation. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Massage treatment, the number of neck-head movements in the Treat0to5 period was significantly different to the Post10to15 period (p = 0.037). IQR = Interquartile range. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

Forelimb movements

The number of forelimb movements per period for each treatment is presented in Figure 10. For forelimb movements during the Massage treatment, there was a significant overall effect of period (p = 0.044). No significant differences between treatments were found during post-hoc pairwise testing. For forelimb movements during the Clicker treatment, there was a significant overall effect of period (p = 0.015). A significant difference between the Post5to10 and Post10to15 periods (p = 0.023), with a median and range of 47 (6-383) and 23 (0-231) forelimb movements respectively, was found during post-hoc pairwise testing.

For forelimb movements during the other treatments, there was no significant overall effect of period (Control p = 0.505, Scrub p = 0.407).

Within each period, the number of forelimb movements did not vary significantly by treatment (Pre-5to0 p = 0.640, Treat0to5 p = 0.299, Post5to10 p = 0.087, Post10to15 p = 0.152).

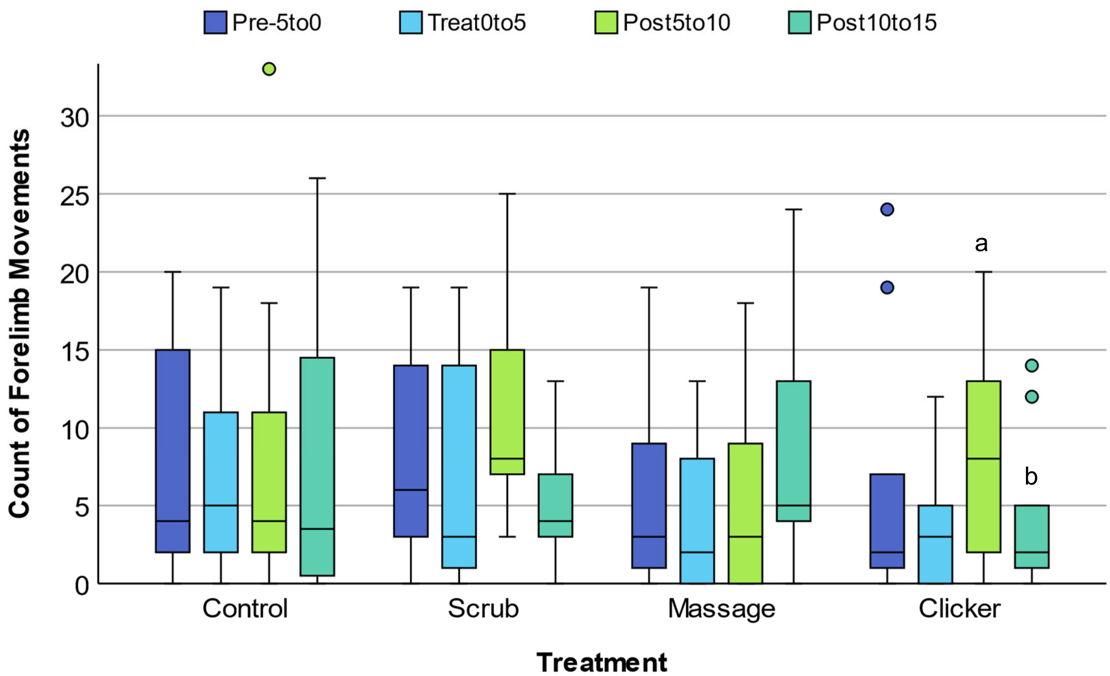


Figure 10 Boxplot of forelimb movements by treatment, clustered by period. Circles indicate potential outliers (1.5-3*IQR), and asterisks indicate extreme outliers (>3*IQR). Some outliers have been removed to aid visualisation. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Clicker treatment, the number of forelimb movements in the Post5to10 period was significantly different to the Post10to15 period (p = 0.023). IQR = Interquartile range. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

Hindlimb movements

The number of hindlimb movements per period for each treatment is presented in Figure 11. For hindlimb movements during the Massage treatment, there was a significant overall effect of period (p = 0.003). A significant difference between the Treat0to5 and Post10to15 periods (p = 0.003), with a median and a range of 3 (0-11) and 11 (1-28) hindlimb movements respectively, was found during post-hoc pairwise testing.

For hindlimb movements during the other treatments, there was no significant overall effect of period (Control p = 0.253, Scrub p = 0.221, Clicker p = 0.194).

Within each period, the number of hindlimb movements did not vary significantly by treatment (Pre-5to0 p = 0.571, Treat0to5 p = 0.218, Post5to10 p = 0.771, Post10to15 p = 0.412).

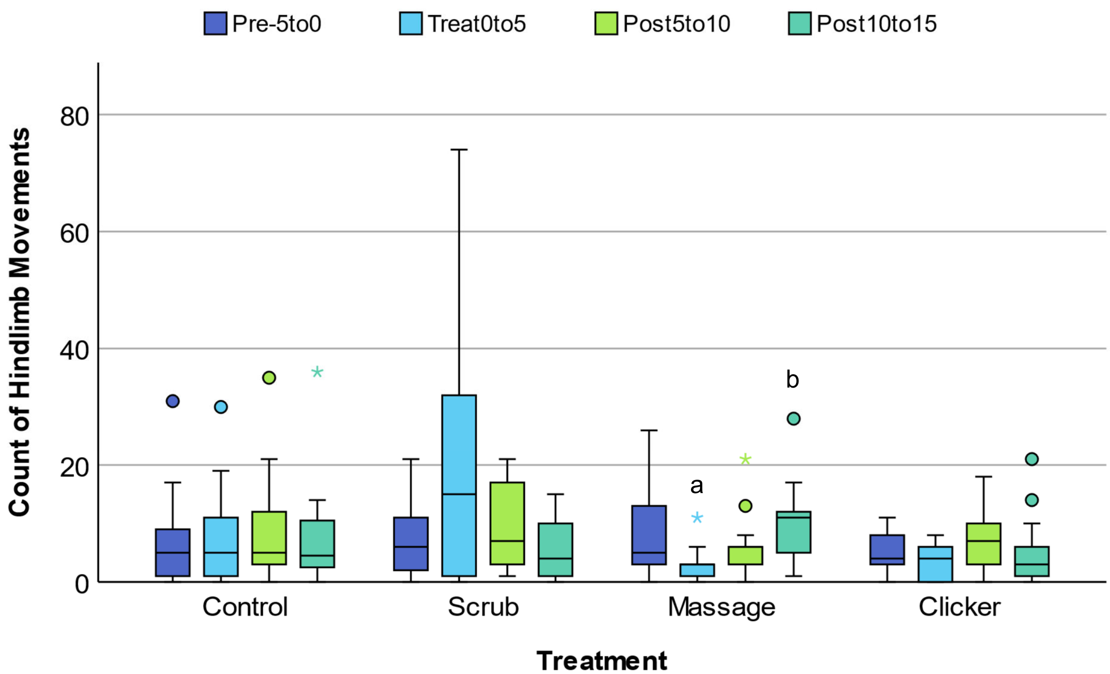


Figure 11 Boxplot of hindlimb movements by treatment, clustered by period. Circles indicate potential outliers (1.5-3*IQR), and asterisks indicate extreme outliers (>3*IQR). Some outliers have been removed to aid visualisation. For analysis of period within each treatment, differing letters indicate a significant difference between means. Within the Massage treatment, the number of hindlimb movements in the Treat0to5 period was significantly different to the Post10to15 period (p = 0.003). IQR = Interquartile range. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

Tail movements

The number of tail movements per period for each treatment is presented in Figure 12. Within each treatment, the number of tail movements did not vary significantly by period (Control p = 0.741, Scrub p = 0.331, Massage p = 0.542, Clicker p = 0.267).

For tail movements during the Treat0to5 period, there was a significant overall effect of treatment (p = 0.020). A significant difference between the Massage and Scrub treatments (p = 0.023), with a median and a range of 0 (0-30) and 4 (0-52) tail movements respectively, was found during post-hoc pairwise testing.

For tail movements during the Post5to10 period, there was a significant overall effect of treatment (p = 0.021). No significant differences between treatments were found during post-hoc pairwise testing. For tail movements during the other periods, there was no significant overall effect of treatment (Pre-5to0 p = 0.392, Post10to15 p = 0.438).

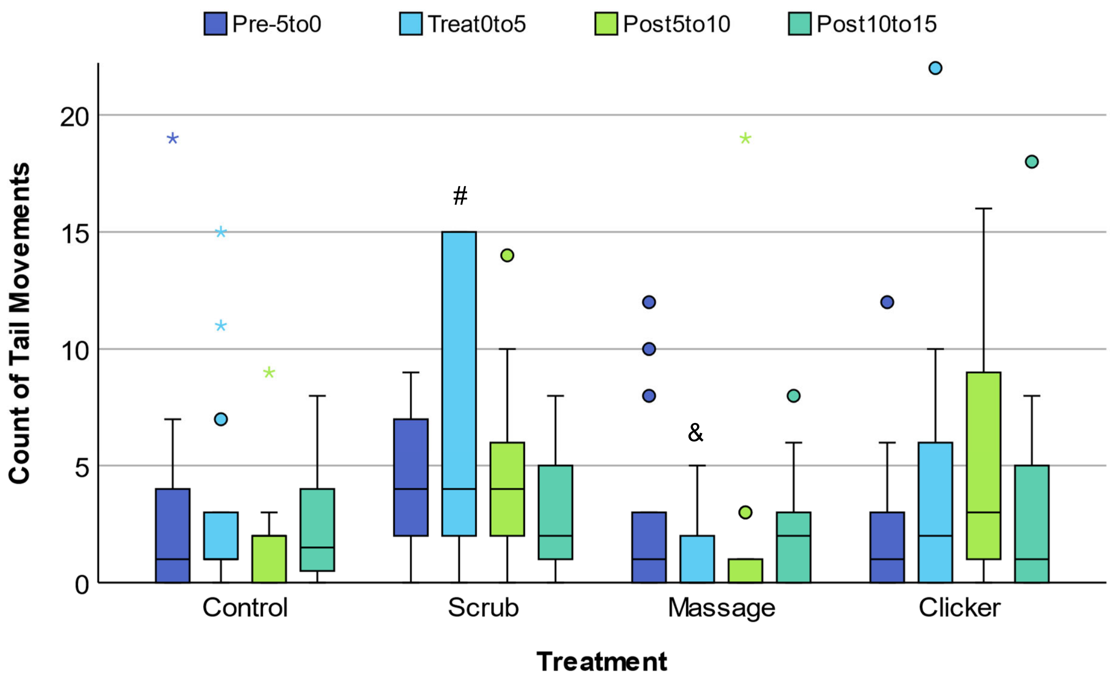


Figure 12 Boxplot of tail movements by treatment, clustered by period. Circles indicate potential outliers (1.5-3*IQR), and asterisks indicate extreme outliers (>3*IQR). Some outliers have been removed to aid visualisation. There were no significant differences between means for period within each treatment. For analysis of treatment within each period, differing symbols indicate a significant difference between means. Within the Treat0to5 period, the number of tail movements in the Scrub treatment was significantly different to the Massage treatment (p = 0.023). IQR = Interquartile range. Pre-5to0 = five-minute period preceding application of treatment, Treat0to5 = five-minute period of treatment application, Post5to10 = five-minute period after treatment application has ended, Post10to15 = five-minute period following Post5to10.

Urination and Defecation

Urination and defecation did not occur often enough to be analysable.

4.4 Discussion

Research on wither massage suggests the treatment would have a pleasant, relaxing effect, and the inguinal scrub treatment would have an aversive, arousing effect on all horses (Feh et al. 1993; McBride et al. 2004; Janczarek et al. 2018b). It was also thought that the Clicker treatment might induce initial positive emotion due to anticipation of the food treat, and then negative emotion as the food was not forthcoming (Willson et al. 2017). However, in this trial, the results showed a large variation in magnitude and direction of change, which may be due either to lack of sensitivity of the measurements or biological variation in the subjects. The results were difficult to interpret and did not show many significant differences between the putatively positive and putatively negative emotion-inducing stimuli. The variability of the data may have overshadowed any real differences between treatments and/or periods.

Contrary to expectations, massage of the wither area, the putatively positive emotion inducing experience, did not result in a significant decrease in heart rate. Similarly, the putatively negative Scrub treatment did not result in a significant increase in heart rate. It was thought that the relative humidity above the skin might increase during the negative condition, due to an increase in sweat production. Conversely, the relative humidity above the skin was not affected by treatment or period. Nonetheless, this reflects a finding in sheep where neither body surface temperature nor body surface humidity varied with emotional valence (Reefmann et al. 2009a).

Horses differed in their reactions to the treatments. The design of this experiment applied the same treatments to all horses with the assumption that they would be perceived as pleasant and unpleasant, and induce positive and negative emotion, respectively. This led to two fundamental issues. First, this assumption did not necessarily hold up. Some horses appeared to respond in a paradoxical way to the treatments. It appeared that some reacted negatively to the presumed positive Massage treatment, and some reacted positively to the presumed negative Scrub treatment. Individual differences may be due to differences in genetics and/or previous experience of the stimuli. This highlights a problem with using universal presumed positive and negative experiences, and underlies one of the fundamental problems when researching emotion in animals. Whether the animals actually perceive the stimulus with the same valence we presume it will induce is unknown. What horses like cannot necessarily be assumed and generalised across a population.

The second issue involved a lack of a gold standard method for identification of emotional valence. This can lead to a circular argument whereby the valence of the emotion being measured is identified by subjective inference. This experiment highlighted the problem that a researcher may set out to measure positive emotion, but how can it be known when the animal is experiencing this emotion, and that they are measuring what they purport to measure? A problem common in other research (Kremer et al. 2020). A mechanism for verifying the presumed valence on an individual basis is required in future research.

Confounded temperature measures were a further issue. Mean eye temperature and skin temperature increased over time with the Control treatment. There may be confounding effects of ambient temperature over time. Ambient temperature should be accounted for during analysis in future research.

Despite these methodological flaws, some promising results were seen. While the number of tail movements did not vary across periods within each treatment, there were more movements during application of the Scrub treatment than the Massage treatment. This may reflect differences in arousal level or agitation between the two treatments and warrants further investigation.

Several changes were seen in the post-treatment periods. Within the Scrub treatment, the RMSSD/SDNN ratio measure of heart rate variability increased in the period following treatment. This may reflect a change in the balance of the autonomic nervous system, with an increase of the parasympathetic nervous system relative to the sympathetic nervous system towards a level similar to that seen prior to the treatment application, which may be due to recovery from the treatment. Measurement of heart rate variability during and post treatment warrants further investigation of whether it may reflect emotional arousal or valence.

The median number of neck-head movements increased by almost three-fold, and the median number of hindleg movements was more than three-fold higher in the second five minutes following application of the Massage treatment compared to during application. It may be that horses moved their heads up and down and altered hind limb position more in agitation or frustration following ending of a desirable treatment, and that head and hindleg movements reflect negative valence and/or increased arousal. This suggestion is supported may by the literature in sheep and cattle (Hemsworth et al. 2011).

In the first five minutes following application of the Clicker treatment, the median number of forelimb movements was more than double that during the second five-minute period after treatment. This may reflect agitation or frustration due to an unmet expectation of receipt of a food treat (Willson et al. 2017), which was not likely a component of the other treatments, and possibly negative valence and/or increased arousal.

Even so, aggregation of movements by body part may not be ethologically relevant, for example, resting a hindlimb has a different ethological meaning than a hind kick, a head move to the high position is different to a head in a mid or low position (McDonnell 2003; McGreevy 2012). On the other hand, perhaps a more aroused horse moves more, and the count of moves reflects arousal level rather than emotional valence. However, in this study, the overall movement count was not found to be a useful metric for differentiating groups during preliminary data analysis.

The effect of missing data for heart rate variability and respiratory rate is unknown but considered likely to be minor due to the relatively small deficit. However, the skin temperature and humidity results may have been more affected by the larger amount of data missing. Caution is advised in interpretation. The findings from this study were severely limited due to design flaws around the assumption of valence. Although some significant results were found, interpretation of their meaning with respect to emotion is likely of limited use.

4.4.1 Conclusion

The interpretation of the results of this trial in relation to the subjective component of emotion is difficult. Critique of this experiment revealed that base assumptions about emotional valence were invalid, and it highlighted an important circular argument relating to verification of emotional valence. The results of this experiment are a stimulus for further work.

Future studies will need to address these issues by tailoring emotion inducing experiences to the individual animal, and objectively verifying emotional valence independent from the potential indicators that are under investigation. Control of confounding factors, such as ambient temperature, and avoiding the aggregation of ethologically disparate behaviours, may be useful. Separation of the effects of arousal and valence is also challenging.

Chapter 5 Application of a novel method of preference testing to enable determination of a horse's most preferred emotional experience

5.1 Introduction

A problem with research on emotion in animals is the assumption that a particular stimulus will induce a particular emotion (Lawrence et al. 2019). A preference for a particular stimulus suggests it is pleasurable and induces a more positive emotion than the alternatives (Humphrey 1972; Boissy et al. 2007b; Dawkins 2008). This experiment was designed to address this issue by using preference testing to identify the stimulus that is most preferred by an individual horse, for use in subsequent investigations of indicators of emotion.

Preference and aversion test methods have been traditionally used to determine which aspects of an environmental a species likes, or dislikes, relative to a reference resource or stimulus (Appleby et al. 2011). Preference tests provide information about relative motivations to obtain or avoid stimuli and may rank them relative to each other (Kirkden et al. 2006).

Experiments designed to establish preference often utilise pairwise presentation of options via a Y- or T-maze, or open access, where an animal can access both options and choose how much time to spend with each (Kirkden et al. 2006; Grandin 2015). The preferred option is the one that the animal chooses more often, consumes more of, or spends more time with. In these kinds of preference tests, the total number of trials required for assessing preferences for a range of stimuli can be quite large. For example, with four stimuli, there would be 12 pairwise combinations to present to each animal. After allowing for crossover of left and right maze arm to counteract effects of a lateral motor or location bias, the total would be 24 combinations.

An alternative approach may be to allow free choice of four simultaneously available locations that have been associated with the delivery of one of four alternative stimuli. Provided that the subjects can discriminate, recall, and reliably choose between four stimuli based on cues and/or spatial location, the time spent in each location may indicate a relative preference for the associated stimulus. This approach, coined the 'location-associated simultaneous preference test' by the author, may reduce the time and costs.

Previous research has been performed in horses to examine preferences for some resources and situations, for example: bedding (Hunter et al. 1989; Mills et al. 2000; Pfister et al. 2017), social contact (Søndergaard et al. 2011), stallions (Burger et al. 2018), orientation during transport (Smith et al. 1994; Gibbs et al. 1999), thermal comfort (Holcomb et al. 2014; Holcomb et al. 2016; Jorgensen et al. 2016; Mejdell et al. 2016; Jorgensen et al. 2019; Mejdell et al. 2019), lateralisation (Austin et al. 2012; Siniscalchi et al. 2014; Inoue et al. 2019), exercise (Lee et al. 2011), being ridden (König von Borstel et al. 2012), human postures (Smith et al. 2018), water bowls (Krawczel et al. 2006), and many parameters of food (O'Connor et al. 2004; Goodwin et al. 2005; Muller et al. 2007; Allen et al. 2013; Randall et al. 2014; van den Berg et al. 2016; DeBoer et al. 2017; Grey et al. 2017; Heuschele et al. 2018; Janczarek et al. 2018a; Catalano et al. 2019). Some methodological aspects of those experiments are pertinent, for example, the duration and repetition of exposures, familiarity with treatments, personnel, equipment and facilities, and nature of cues. However, the current experiment is unique in purpose, design, and the presented options.

Briefly, generally important factors in preference and aversion testing fall into three areas: study design, within subject factors, and extraneous/environmental factors (Hunter et al. 1989; Kirkden et al. 2006; Fraser et al. 2018). Design factors include the type of research question being asked (including motivation to obtain or avoid a single resource; preference amongst resources; strength of preference/motivation), biological significance of the choices and immediacy of survival benefit, and the species' motivation to perform a behaviour or obtain the outcome of that behaviour.

Within subject factors include previous experience of the stimuli, satiety and satiation rate, emotional state, and diurnal or seasonal variation. Extraneous/environmental factors include social isolation, effect of one stimulus over another, and concurrent availability or exclusivity of the stimuli and other resources. Additional factors affecting learning and decision making may also be important such as individual temperament (fearfulness, reactivity, sensory sensitivity), environmental distractions (noise, movement), age, breed, sex, psychological stress, arousal level, motor laterality, and attention span (Sappington et al. 1997; Lansade et al. 2010).

The aims of the current experiment were to identify a set of individual horses that displayed a reliable preference for one of the offered stimuli using the 'location-associated simultaneous preference test', to determine their most preferred stimulus, and to confirm that the putatively aversive stimulus was not preferred by them. Two methods of assessing the preference data were compared to see if preference could be determined without the use of arbitrary criteria, and if strength of preference could be assessed.

The stimuli likely to be preferred or avoided (and thus induce positive or negative emotion) were chosen based on previous research. Grooming or massage of the withers region may mimic allogrooming and was expected to be the most desirable stimulus, as it is a common site of allogrooming in horses, has been associated with positive behavioural and/or physiological responses, has been used as a positive reinforcer, is suggested to be pleasant, and may be of biological significance (Feh et al. 1993; McBride et al. 2004; Reefmann et al. 2009a; McGreevy 2012; Scopa et al. 2020). The presence of a non-interacting person was expected to be desirable for some horses, perhaps due to safety, or a previous association with food, and the absence of an additional stimulus (no interference) desirable for other horses (Fureix et al. 2009; Smith et al. 2018; Scopa et al. 2020). The spray stimulus, a remotely activated lightly scented aerosol sprayed on the girth area, was predicted to be mildly to moderately aversive for all horses, perhaps inducing fear, and has been used in studies of avoidance task learning, temperament, and reactivity in horses (Seaman et al. 2002; Visser et al. 2003; Lansade et al. 2010) and may stimulate a startle response. It was anticipated that horses would choose to spend more time in locations where they expected to receive their preferred stimulus, and would avoid locations where they expected an aversive stimulus to occur (Boissy et al. 2007b; Fraser et al. 2018).

The results from this experiment were used to inform subsequent investigations of indicators emotional experience in horses, and to determine if the research approach may be appropriate for future research on indicators of emotion in animals.

5.2 Materials and methods

All procedures were conducted in accordance with the Massey University Code of Ethics for Animals (MUAEC approval 10/92).

5.2.1 Animals and housing

Twenty horses from a teaching herd maintained at the Massey University Veterinary Large Animal Teaching Unit, Manawatu, New Zealand, were used. Due to the novelty of the experiment, an *a priori* power analysis could not be performed. The number of subjects was estimated based on the theorised subjects required for the following experiment, allowing for losses, while balancing ethical and practical constraints (Taborsky 2010).

Throughout the experiment, horses were kept at pasture in established management groups so that social bonds were not disrupted. On the day of testing, horses were kept for the day in holding pens at the experimental facility. Horses were accustomed to being handled. Their diet consisted of grass and small amounts of hay with water provided *ad libitum*. Ages ranged from 5-22 years old. There were three geldings, 16 mares, and one ovariectomised mare. Twelve of the horses were Standardbred and eight were Thoroughbred or Thoroughbred cross breeds (Table 4).

Table 4 Horse age, sex, colour, breed, and group number. Horses were kept, trained, and tested in groups in the order shown.

Group	ID	Name	Sex	Age (years)	Breed	Colour
1	1	Shane	G	6	SB	Brown
	2	Mr DG	G	13	SB	Bay
	3	Lucy	M	10	SB	Bay
	4	Spice	M	5	SB	Bay
2	5	Semitone	M	22	SB	Bay
	6	Phoenix	M	11	TB	Bay
	7	Hattie	M	8	TB	Bay
	8	Streisand	M	15	SB	Bay
3	9	Frankie	M	5	SB	Black
	10	Minty	M	14	SB	Bay
	11	Chonty	M	9	SB	Bay
	12	Whitney	M	17	TB	Chestnut
4	13	Laurel	M	11	TB	Bay
	14	Pixie Fae	M	13	SB	Bay
	15	Toy	M	22	SB	Bay

Group	ID	Name	Sex	Age (years)	Breed	Colour
	16	Grace	M*	17	TBx	Chestnut
5	17	32	M	15	TB	Bay
	18	Zoe	M	16	TB	Bay
	19	Daisy	M	15	TB	Bay
	20	Fatty	G	11	SB	Bay

ID = Identification number, G = Gelding, M = Mare, * = Ovariectomised, SB = Standardbred, TB = Thoroughbred, x = Cross

5.2.2 Facility setup

The experiment was conducted in a paddock at the Massey University Veterinary Large Animal Teaching Unit. An outdoor semi-circular arena (20m diameter) with a start box (2m x 3m) was constructed using posts (1.8m high) and three strands of white fencing tape (2.5cm wide) (Figure 13). There was an entry/exit gate, made from tape, into the start box, and a single horizontal rail that acted as a gate between the start box and arena. The grass in the arena was sprayed with herbicide and closely mown to reduce any effect of feed availability on the horses' location preference. The arena was divided into four sectors (numbered one to four), plus a semi-circular area around the start box gate (sector 'o'), using white paint. Sector 'o' was included as a no-choice area, as it was difficult to determine which of the four numbered sectors a horse was in when close to the start box. Holding pens made from temporary fencing were erected around the outside of the semi-circular arena to contain companion horses so that the test horse was not visually isolated from conspecifics, and the proximity to another horse was equal for each sector. Water and grass were available *ad libitum* to companion horses in the holding pens.

Seven video cameras were set up around the semi-circular arena to record the location of the test horse and its behaviour. Vertical and horizontal scales were provided by markings on vertical posts and on the ground.

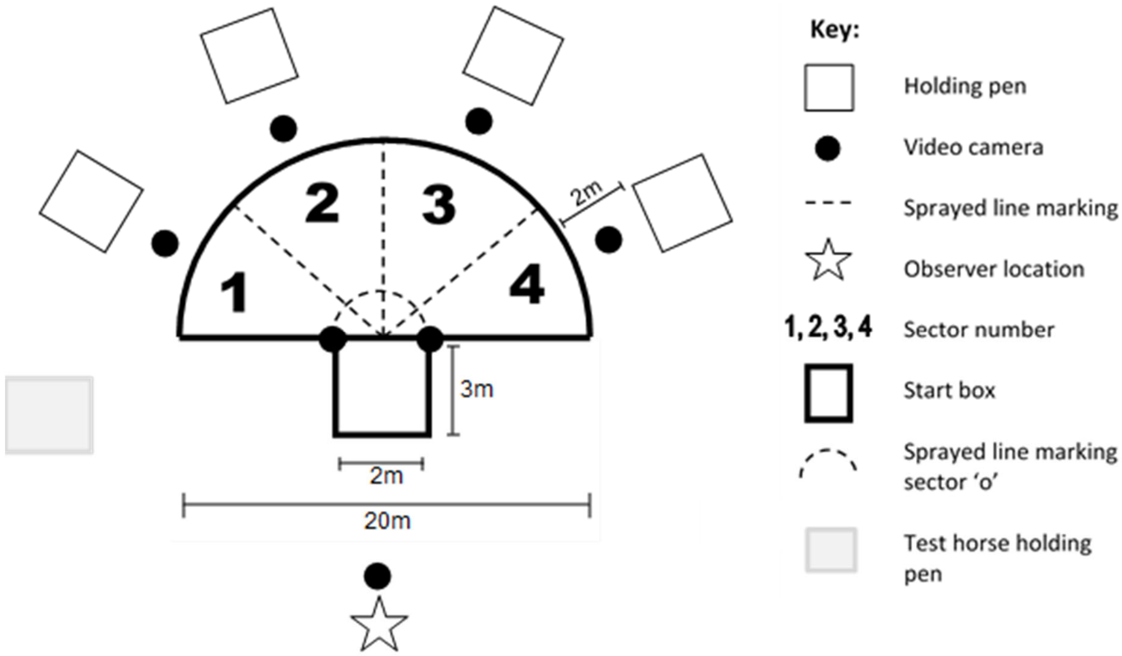
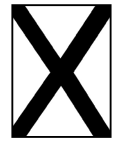
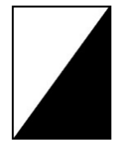
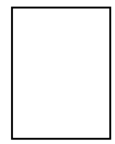
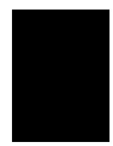


Figure 13 The outdoor, uncovered experimental facility consisted of a 20 m semi-circular arena with a 2 m x 3 m start box, was formed from posts and fencing tape.
The grass in the arena was killed and closely mown then marked with white paint to divide the semi-circle into four sectors and a semi-circular area by the start box (sector 'o'). Holding pens, 4 m x 5 m, were constructed from temporary fencing.

5.2.3 Experimental procedure

Following habituation and training, a horse's preference among four stimuli (Table 5) was determined by comparing the proportion of time it spent in locations that had previously been associated with the experience of those stimuli, which included a black and/or white visual cue. Two people were used to impose the treatments, the person delivering each stimulus was consistent for each horse (but differed between horses).

Table 5 Experimental stimuli and associated black and/or white visual cue.

Stimulus	Description	Visual Cue
Person (P)	A previously unfamiliar person stood motionless in the sector, with a relaxed posture, not interacting or making eye contact with the horse. The horse could choose to interact with the person and was able to sniff or lick them (horse-initiated contact). The person stepped back out of reach if the horse tried to bite. The same person was used for each repetition for an individual horse.	
Groom (G)	A previously unfamiliar person (different to P) stood in the sector. Once the horse entered the sector, they approached the horse and groomed the left wither area using a rubber grooming tool in a circular motion (human-initiated contact). The same person was used for each repetition for an individual horse.	
Spray (S)	A single, lightly scented aerosol spray was emitted after 10 seconds, and repeated once every 20 seconds, for a total of 4 sprays in 90 seconds, from a remote-controlled spray unit (Petsafe Spray Remote Trainer System SRT-200, Radio Systems Corporation, Knoxville, USA) located behind the right elbow. The spray device was placed under a surcingle on the right side of the horse's chest, midway between ventrum and dorsum, and was worn during all sessions.	
No-stimulus (C)	Nothing was present in the sector except the visual cue	

Horses were in the same groups of four for each part of the experiment. Each group was habituated, trained, and tested over a five-day period in Summer. At the start of each day, a group of horses, plus a companion horse, were led to the experimental facility and randomly put into the holding pens. While in the holding pen, a heart rate monitor (Polar Equine S810iTM, Polar Electro Oy, Helsinki, Finland) and the spray device (Petsafe Spray Remote Trainer System SRT-200, Radio Systems Corporation, Knoxville, USA) were placed under an elastic surcingle, with the spray device on the right side of the horse's chest, midway between ventrum and dorsum. During the day horses were moved between holding pens to reduce risk of systematic bias in location preference. The horses were led back to their paddock at the conclusion of the day's events. The schedule of events for each horse is given in Table 6 with the events described in the following sections.

Table 6 Experiment schedule of events for each horse.

Day	Event
1	Habituation sessions (up to 3) Training session 1
2	Training session 2 Training session 3
3	Training session 4 Training session 5
4	Preference test 1 Training session 6 Preference test 2 Training session 7
5	Preference test 3 Training session 8 Preference test 4

Habituation sessions

Horses were acclimated to the personnel, facilities, and equipment, over two or three habituation sessions, as follows. The handler led the horse into the start box, closed the gate, and returned to the observer location. After three minutes, the handler opened the gate between the start box and arena. The horse entered the arena of its own volition, the gate was closed, and the handler returned to the observer location. The horse was free to investigate the whole semi-circular arena for 10 minutes. During this time, the horse's behaviour was monitored, and the time taken to achieve habituation criteria (Table 7) was recorded. The horse was then returned to a holding pen whilst the other horses in the group were habituated. The process was repeated in further sessions so that the habituation criteria were met in at least two sessions. The length of the habituation sessions was based on pilot observations plus a generous margin.

Care was taken with handling within and outside of the experimental setting to avoid horses inadvertently forming negative associations with personnel by maintaining calm, quiet handling by familiar handlers, and avoiding the use of punishment.

Table 7 Criteria for habituation to facilities and equipment during habituation sessions.

Category	Behaviour criteria
Head position	Mostly neutral (normal standing angle) to low (head below wither height)
Ears	Relaxed, not swivelling or orientated to a particular stimulus
Mouth	Relaxed, not tight
Vocalisation	None
Movement	None or slow walking
Investigation	Grazing/investigating ground or resting continuously for ≥ 10 sec

Training

Horses were trained to associate each stimulus with a spatial location (sector) and a visual cue. Visual cues (Table 5) were painted on a 60 cm x 100 cm plywood board, using black and/or white paint, and one was displayed in each sector. Although spatial cues are thought to be more important to horses than visual cues, visual cues were included to reinforce the spatial cue, to help horses learn the association between location and stimulus, for identification on video, and to assist conduct of the experiment (Martin et al. 2006). The nature of the visual cues was based on previous research that suggested that horses were able to discriminate black and white symbols, and that presentation at ground level was better than an elevated position (Hall et al. 2003; Hanggi 2003; Hanggi et al. 2007; Christensen et al. 2008; Hanggi et al. 2009a; Hanggi 2010).

During training sessions, horses were exposed to each stimulus a total of eight times over several days so that learning was enhanced and memory consolidation could occur (Arnold et al. 2008; Demant et al. 2011). The number of training trials to conduct for a free choice preference test seems to be largely arbitrary. For example, a preference test of stall confinement time in pigs allowed for 12 exposures to each confinement period over six days (Spinka et al. 1998). The authors suggested less training was required in future. A preference test on horses involving an operant task for release from confinement found that on average one 20 minute training session on three consecutive days was sufficient for learning to push a panel (Lee et al. 2011).

The order in which the horses were initially trained and tested was selected at random. To control for time-of-day effects, this order was the same each day, so that the approximate time of day was generally consistent for each horse. Stimuli were presented in the same order for all horses (Person, Groom, Spray, No-stimulus) to reduce potential carryover effects of the Spray stimulus. If the Spray stimulus was presented last, it was thought that it may impact on the horse's propensity to enter the arena on subsequent occasions due to a negative association. Therefore, the No-stimulus condition was presented last to allow the horse time to recover before leaving the arena. The sector associated with each stimulus (stimulus location) was initially randomly allocated for each horse using a modified Latin square design of the 24 possible stimulus-by-sector combinations. The stimulus location was then consistent for each individual horse across all training and testing sessions. The number of times each stimulus was allocated to each sector location is presented in Table 8.

Table 8 Distribution of stimuli across sector locations following random allocation. Values are the number of times that each stimulus was allocated to each sector across the 20 horses.

Stimulus	Sector Location				Total
	1	2	3	4	
Person	4	5	7	4	20
Groom	6	5	4	5	20
Spray	5	7	3	5	20
No-stimulus	5	3	6	6	20
Total	20	20	20	20	

Each stimulus was presented to each horse once during each training session. The four visual cues were positioned in the relevant sectors for the horse. The horse was led into the start box and the handler returned to the observer location. After one minute, the person applying the stimulus (if applicable) stood next to the visual cue in the relevant sector and the handler opened the gate into the arena. If a horse did not enter the arena after 30 seconds, it was encouraged using the following stimuli sequentially: finger clicks, hand claps, patting the rump, then being led in. The gate was closed after the horse entered the arena, and the handler returned to the observer location. For training sessions only, the sectors were fenced off with fencing tape, so that the horse could only enter the one relevant sector when released from the start box.

The stimulus was presented for four minutes, except for the Spray stimulus, which lasted for 1.5 minutes. The Spray was given at 20 second intervals, starting after 10 seconds had elapsed (to allow time for the handler to return to the observer location), for a total of four sprays. After the stimulus presentation period, the horse was returned to the holding pen. Following an interval of at least 10 minutes, and provided the horse's HR was ≤ 50 bpm, the procedure was repeated for each stimulus in order (Person, Groom, Spray, No-stimulus), in the relevant location for each horse.

Preference testing

For the test sessions on days four and five, the arena was set up as for the training sessions, except sectors were not fenced off. To accustom the horse to being able to freely move between sectors, each horse was led once through the arena immediately before each test session following a consistent route.

After one minute in the start box, the horse was released into the arena for two minutes, before being returned to a holding pen. The handler returned to the observer location when not handling the horse. Although visual cues were present, no stimuli were presented during test sessions to avoid any influence on the horse's choices by the presence of people in the arena. This meant that preferences were assessed based on the learned association between the visual and spatial cues and the stimuli. The association was reinforced by conducting a single training session prior to each test. The test duration was a compromise between allowing time to explore the areas, with knowledge that the first choice an animal makes does not necessarily reflect its preference, and risk of dissociation of location with stimulus after an unknown length of time (Kirkden et al. 2006).

5.2.4 Measures

Horse behaviour was recorded using digital video cameras (Sony DCR-SR85E, Auckland, New Zealand). The time of day when the test horse entered the start box, was released from the start box, and entered and exited the arena were recorded. The number of sessions and time taken to achieve habituation criteria were recorded. The time spent in each sector, including the 'o' sector, during habituation and test sessions, was measured from video analysis and used to determine the most preferred stimulus for each horse. A horse was deemed to enter a sector when its surcingle crossed the painted line. The horses that were in the holding pens, the weather conditions, and any extraneous noises or events were also recorded.

5.2.5 Data handling and analysis

An individual horse's preference was considered in terms of its choice of where it spent the most time during each test, and averaged across all tests. For each horse, the percentage of time spent in each sector (1-4) during each of the four tests, the mean time, standard error (SE), and median time across all tests was calculated.

The stimulus associated with the location where a horse spent the highest percentage of time during a test was determined to be the horse's preferred stimulus for that test. The 'most preferred stimulus' for each horse was determined as the stimulus that was most often preferred across all tests i.e., the stimulus location that was preferred in the most tests.

Each horse's preference was then assessed in two ways to determine if arbitrary criteria could be replaced with equation-based assessment.

Criterion method

The Criterion method combined the number of times that a stimulus-location was preferred across tests with the average percentage of time spent in the stimulus-location to give measures of preference consistency and strength. Collectively, the preference consistency and strength were used to determine preference reliability, based on arbitrary cut off values.

'Preference consistency' was defined as the percentage of tests that the most favoured stimulus was preferred in, with ≥ 51% arbitrarily considered as good consistency (i.e., the same stimulus was preferred in the majority of tests) (Smith et al. 2018). 'Preference strength' was defined as the mean proportion of time across all tests that a horse spent in the stimulus location for the most preferred stimulus. Arbitrary cut offs were used to interpret the preference strength: Strong if ≥ 70%, Moderate if 50-69%, or Weak if < 50%. An individual horse was deemed to have displayed a 'Reliable preference' if the consistency was at least 51% and the preference strength was at least 70%.

Using a definition of preference as a horse choosing the same sector in at least two of three or three of four tests (i.e., majority of tests), and a definition of no-preference as a horse choosing a different sector in at least two tests, the chances of a horse falsely displaying a preference and falsely displaying no-preference were calculated. A multinomial exact goodness-of-fit test was used to determine if the distribution of horse preferences differed from the expectation of random distribution (Smith et al. 2018).

Index method

The second method of assessing preference utilised calculation of a choice index for each stimulus (Beausoleil 2006) and a similarly calculated time index, and was termed the index method. Choice and Time Index methods were explored to determine if they would obviate the need for use of arbitrary cut offs as in the Criterion method. The Choice index was based on the number of times a treatment location was preferred, and the Time index was based on the mean time spent in each location across all tests.

A positive index was deemed to indicate a preference for that stimulus, and the magnitude of the index described the relative preference strength. The equations are given below for the Person stimulus. Note that the equations given below for calculation of choice index and time index, are for Person (P) as examples and would change accordingly for other stimulus calculations.

Equation for the calculation of choice index for the Person stimulus:

$$\text{Choice Index P} = \frac{[\text{Choices P} - (\text{Choices G} + \text{S} + \text{C})]}{(\text{Choices P} + \text{G} + \text{S} + \text{C})} \times 100$$

Where P = Person stimulus, G = Groom stimulus, S = Spray stimulus, C = No-stimulus, and Choices is the number of tests in which the stimulus was the preferred one.

Equation for calculation of time index for the Person stimulus:

$$\text{Time Index P} = \frac{[\text{Time P} - (\text{Time G} + \text{S} + \text{C})]}{(\text{Time P} + \text{G} + \text{S} + \text{C})} \times 100$$

Where P = Person stimulus, G = Groom stimulus, S = Spray stimulus, C = No-stimulus, and Time is the mean percentage of time spent in the sector associated with the stimulus across all tests.

Location bias

To investigate potential location preference, the percentage time spent in each sector during each test was summarised by mean and standard error (SE) for each horse. Outliers were identified from assessment of boxplots, checked for errors, and corrected or kept as valid data. Normality was assessed by Shapiro-Wilk test ($p > 0.05$). A series of one-way repeated measures ANOVA (Analysis of Variance) tests were conducted to determine whether there were statistically significant differences in mean time between sectors across all horses, or within those that displayed a reliable preference and those that did not (as determined by the criterion method). Sphericity was assessed using Mauchly's test, and where the assumption was violated, the Greenhouse-Geisser correction was used (Laerd Statistics 2015). The same process was used to determine if there were statistically significant differences in mean time between stimuli locations for the same subsets of horses.

To further check for location bias on an individual basis, an attempt was made to 'normalise' test results for any location preference during habituation. The mean percentage of time spent in each sector during habituation was subtracted from the mean percentage of time spent in the sector during testing. The sector with the highest habituation-normalised mean percentage of time was compared to the sector associated with the 'most preferred stimulus' for each horse for agreement.

Descriptive statistics were produced with Microsoft Excel (version 2107, Microsoft 365, USA). Statistical analyses were performed in IBM SPSS Statistics for Windows (version 27.0, released 2020, IBM Corp., USA) and R: A Language and Environment for Statistical Computing (version 4.2.0, R Core Team, The R Foundation for Statistical Computing, Vienna, Austria). The statistical significance level was set at $p \leq 0.05$. Bonferroni correction of p-values was applied during post-hoc pairwise testing to account for repeated measures.

5.3 Results

The data from one (ID 2, Mr DG) of the 20 horses were excluded from the study due to the horse displaying stereotypic behaviour. The data from another horse (ID 13, Laurel) was excluded as the horse did not engage in any of the tests and spent all its time in the ‘o’ sector.

5.3.1 Habituation

All remaining 18 horses received two 10-minute habituation sessions. The percentage of time that each horse spent in each sector location during each habituation session, with summary statistics for each horse, are presented in the Appendix A.3.1, Table A 1. The mean time to habituation for the first habituation session was 3.75 minutes (SD 2.12, range 1.0-9.0, median 3.25). The mean time to habituation for the second habituation session was 1.69 minutes (SD 0.91, range 0.5-3.5, median 1.5). The mean percentage of time that horses spent in each sector during the two 10-minute habituation sessions is presented in Figure 14.

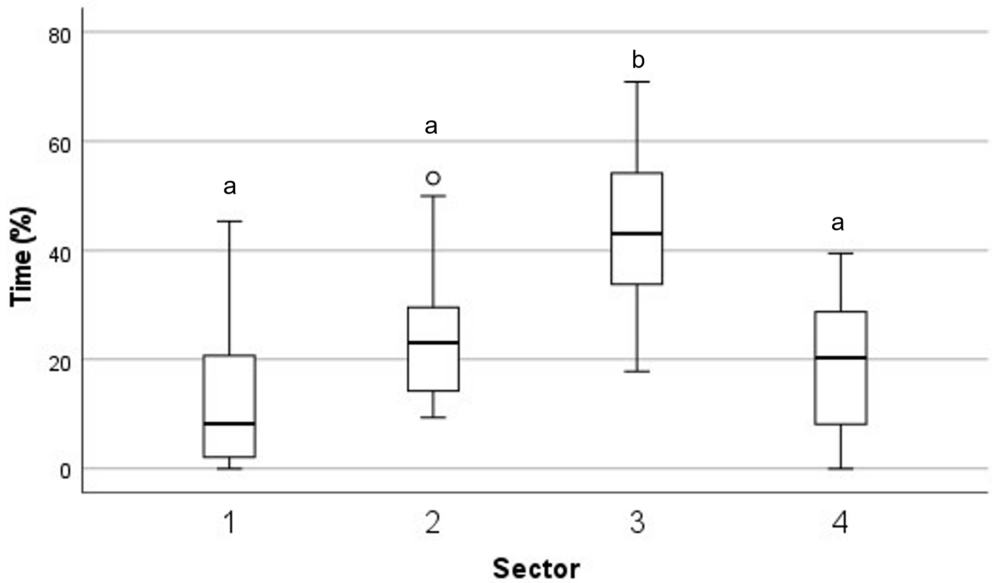


Figure 14 Boxplots of mean time horses spent in each sector during the two 10-minute habituation sessions.
Differing letters indicate a significant difference between means. The mean time spent in sector 3 differed from sectors 1 ($p < 0.001$), 2 ($p = 0.039$), and 4 ($p = 0.001$).

5.3.2 Preference testing

Due to inclement weather, the number of tests for four horses (group 3) were reduced to three. This was accommodated by calculating the preference consistency results as a percentage of the number of tests each horse received.

The percentage of time that each horse spent in each sector location and each stimulus-location during each test, with summary statistics for each horse, are presented in the Appendix A.3.2, Table A 2. From these data, the most preferred stimulus for each horse was determined according to the number of times each stimulus was preferred across the tests, and is presented in Table 10.

5.3.3 Individual preference assessment

Criterion method

The consistency and strength of preference for each horse as determined by the criterion method are presented in Table 10. Based on the preference reliability criterion, ten of the 18 horses (55.6%) displayed a reliable preference. Eight horses did not display a reliable preference (44.4%), although one of them (ID 10, Minty) displayed a consistent (consistency 66.7%), but weak preference (strength 46.4%).

Of the horses that displayed a reliable preference, 6/10 (60.0%) preferred the person stimulus, 3/10 (30.0%) preferred the grooming stimulus, and 1/10 (10.0%) preferred the no-stimulus situation. For horses that displayed a reliable preference, the spray stimulus was not preferred in any tests.

The chances of horses falsely displaying a preference or absence of preference are presented in Table 9.

Table 9 The probability of a horse falsely displaying a preference or absence of preference.

Error	Three tests	Four tests
Falsely display a preference	6.25% (=0.25^2*100)	1.6% (=0.25^3*100)
Falsely display no-preference	37.5% (=1*0.75*0.5*100)	9.4% (=1*0.75*0.5*0.25*100)

A multinomial exact goodness-of-fit test showed that the distribution of horse preferences (Person, N = 6; Groom, N = 3; Spray, N = 0, No-stimulus, N = 1, no preference, N = 8) differed significantly from the expectation of random distribution (p = 0.008).

Table 10 The number of tests in which each stimulus was preferred based on the percentage of time spent in the sector location, and the most preferred stimulus for each horse.
Preference consistency, preference strength, and preference reliability, as determined using the criterion method.

Group	ID	Name	Number of tests [a]	Preferred stimulus frequency [b]				Stimulus most often preferred	Preference consistency [c] %	Preference strength [d] %	Reliable [e] Y/N
				Groom	Person	Spray	No-stimulus				
1	1	Shane	4	4	0	0	0	Groom	100.0	100.0	Y
	3	Lucy	4	0	4	0	0	Person	100.0	100.0	Y
	4	Spice	4	0	4	0	0	Person	100.0	100.0	Y
2	5	Semitone	4	4	0	0	0	Groom	100.0	88.8	Y
	6	Phoenix	4	1	3	0	0	Person	75.0	79.0	Y
	7	Hattie	4	1	1	2	0	Spray	50.0	49.1	N
	8	Streisand	4	0	2	0	2	Person = No-stimulus	50.0	51.2/47.0	N
3	9	Frankie	3	0	3	0	0	Person	100.0	93.9	Y
	10	Minty	3	0	1	0	2	No-stimulus	66.7	46.4	N
	11	Chonty	3	1	1	1	0	Groom = Person = Spray	33.3	33.9/43.6/20.5	N
	12	Whitney	3	0	3	0	0	Person	100.0	100.0	Y
4	14	Pixie Fae	4	3	1	0	0	Groom	75.0	74.6	Y
	15	Toy	4	1	0	2	1	Spray	50.0	51.3	N
	16	Grace	4	0	3	0	1	Person	75.0	75.5	Y
5	17	32	4	0	1	2	1	Spray	50.0	34.6	N
	18	Zoe	4	0	2	1	1	Person	50.0	43.0	N
	19	Daisy	4	0	0	0	4	No-stimulus	100.0	89.4	Y

Group	ID	Name	Number of tests [a]	Preferred stimulus frequency [b]				Stimulus most often preferred	Preference consistency [c] %	Preference strength [d] %	Reliable [e] Y/N
				Groom	Person	Spray	No-stimulus				
	20	Fatty	4	2	1	0	1	Groom	50.0	48.3	N

ID = Identification number
[a] Horses in group 3 received 3 tests instead of 4, due to prolonged inclement weather
[b] Number of times the horse spent the most time in that stimulus location during tests
[c] Preference consistency was calculated as the percentage of tests in which the most often preferred stimulus was preferred. E.g., for horse ID 6 the most often preferred stimulus was Person, which was preferred in 3 of 4 tests 3/4*100 = 75%. Consistency ≥ 51% was interpreted as good.
[d] Preference strength equals the mean time the horse spent in the most often preferred stimulus location across all tests. Interpretation: Strong ≥ 70%, Moderate 50-69%, Weak < 50%.
[e] A horse's preference was deemed reliable if the preference consistency was ≥ 51% and the preference strength was ≥ 70%. Y = Yes, N = No.

Index method

Choice and Time indices were calculated for each stimulus for each horse, and are presented in Table 11 and Table 12 respectively, along with the preferred stimulus and preference strength, as determined using the index method.

Based on the results of the Choice Index, 11 of the 18 horses (61.1%) displayed a preference. Of those, 6/11 (54.5%) preferred the person stimulus, 3/11 (27.3%) preferred the grooming stimulus, and 2/11 (18.2%) preferred the no-stimulus situation. The Choice Index for the spray stimulus was ≤ 0 for all horses and was -100 for horses that displayed a preference. The preferences, as determined by the Choice Index method, agreed with the 'most preferred' stimulus for each horse that displayed a consistent preference. There was one disagreement with the Criterion method of preference assessment, and that was horse ID 10, Minty. Using the Choice Index, it was found that Minty displayed a preference for the no-stimulus condition, with a strength of 33%, whereas with the Criterion method it was found that, although Minty displayed a consistent preference (66.7%), Minty did not display a reliable preference, as the strength was < 70% (46.4%).

Based on the results of the Time Index 11 of the 18 horses (61.1%) displayed a preference. Of those, 7/11 (63.6%) preferred the person stimulus, 3/11 (27.3%) preferred the grooming stimulus, and 1/11 (9.1%) preferred the no-stimulus situation. The Time Index for the spray stimulus was ≤ 3 for all horses, and ≤ -88 for horses that displayed a preference. For horses that displayed a consistent preference, the preferences as determined by the Time Index method agreed with the 'most preferred' stimulus, except for one. Using the Time Index it was found that horse ID 8, Streisand, displayed a preference for the person condition with a strength of 2%, whereas the Criterion method found that Streisand did not display a consistent (50.0%), or strong (51.2%) preference, and thus not a reliable preference.

Table 11 Choice index for each stimulus for each horse.
The preferred stimulus and preference strength as determined using the Choice index method.

Group	ID	Name	Choice index [a]				Preferred stimulus	Preference strength
			Groom	Person	Spray	No-stimulus		
1	1	Shane	**100**	-100	-100	-100	Groom	100
	3	Lucy	-100	**100**	-100	-100	Person	100
	4	Spice	-100	**100**	-100	-100	Person	100
2	5	Semitone	**100**	-100	-100	-100	Groom	100
	6	Phoenix	-50	**50**	-100	-100	Person	50
	7	Hattie	-50	-50	0	-100	-	-
	8	Streisand	-100	0	-100	0	-	-
3	9	Frankie	-100	**100**	-100	-100	Person	100
	10	Minty	-100	-33	-100	**33**	No-stimulus	33
	11	Chonty	-33	-33	-33	-100	-	-
	12	Whitney	-100	**100**	-100	-100	Person	100
4	14	Pixie Fae	**50**	-50	-100	-100	Groom	50
	15	Toy	-50	-100	0	-50	-	-
	16	Grace	-100	**50**	-100	-50	Person	50
5	17	32	-100	-50	0	-50	-	-
	18	Zoe	-100	0	-50	-50	-	-
	19	Daisy	-100	-100	-100	**100**	No-stimulus	100
	20	Fatty	0	-50	-100	-50	-	-

[a] Positive Choice index values are shown in bold and indicate that the stimulus was chosen/preferred in more than 50% of tests. The magnitude of the index describes the relative preference strength.
- No preference identified

Table 12 Time index for each stimulus for each horse.
The preferred stimulus and preference strength as determined using the Time index method.

Group	ID	Name	Time index [a]				Preferred stimulus	Preference strength
			Groom	Person	Spray	No-stimulus		
1	1	Shane	**100**	-100	-100	-100	Groom	100
	3	Lucy	-100	**100**	-100	-100	Person	100
	4	Spice	-100	**100**	-100	-100	Person	100
2	5	Semitone	**78**	-78	-100	-100	Groom	78
	6	Phoenix	-58	**58**	-100	-100	Person	58
	7	Hattie	-63	-36	-2	-100	-	-
	8	Streisand	-96	**2**	-100	-6	Person	2
3	9	Frankie	-88	**88**	-100	-100	Person	88
	10	Minty	-82	-35	-76	-7	-	-
	11	Chonty	-32	-13	-59	-96	-	-
	12	Whitney	-100	**100**	-100	-100	Person	100
4	14	Pixie Fae	**49**	-49	-100	-100	Groom	49
	15	Toy	-45	-87	-1	-71	-	-
	16	Grace	-100	**51**	-100	-51	Person	51
5	17	32	-91	-46	-31	-33	-	-
	18	Zoe	-83	-14	-58	-45	-	-
	19	Daisy	-100	-90	-88	**79**	No-stimulus	79
	20	Fatty	-3	-47	-100	-50	-	-

[a] Positive Time index values are shown in bold and indicate that the stimulus was preferred based on mean time. The magnitude of the index describes the relative preference strength.
- No preference identified

5.3.4 Location bias

The distribution of stimuli across each sector location is shown in Table 13 for horses that displayed a reliable preference. The number of horses that displayed a reliable preference for each stimulus-sector combination are shown in brackets. Sector 3 was the most frequently preferred location (n = 5).

Table 13 Distribution of stimuli across sector locations for horses that displayed a reliable preference (n = 10).
Values are the number of times that each stimulus was allocated to each sector during initial random allocation. Values in brackets are the number of horses which preferred that stimulus-sector.

Stimulus	**Sector**				**Total**
	1	**2**	**3**	**4**	
Person	2	4 (3)	3 (3)	1	10 (6)
Groom	3	2 (1)	3 (1)	2 (1)	10 (3)
Spray	2	3	1	4	10 (0)
No-stimulus	3	1	3 (1)	3	10 (1)
Total	10 (0)	10 (4)	10 (5)	10 (1)	

To check for potential impact of a location bias on stimulus preference, the time-location and time-stimulus data were analysed across all horses and all tests, plus those with and without a reliable preference (Figure 15 and Figure 16 respectively).

The results of analysis of mean percentage of time across sectors for all horses (All), those that displayed a reliable preference (WithPref), and those that did not (NoPref), are presented in Figure 15. Within each subset of horses, mean time did not vary significantly by sector (All p = 0.165, WithPref p = 0.133, NoPref p = 0.792).

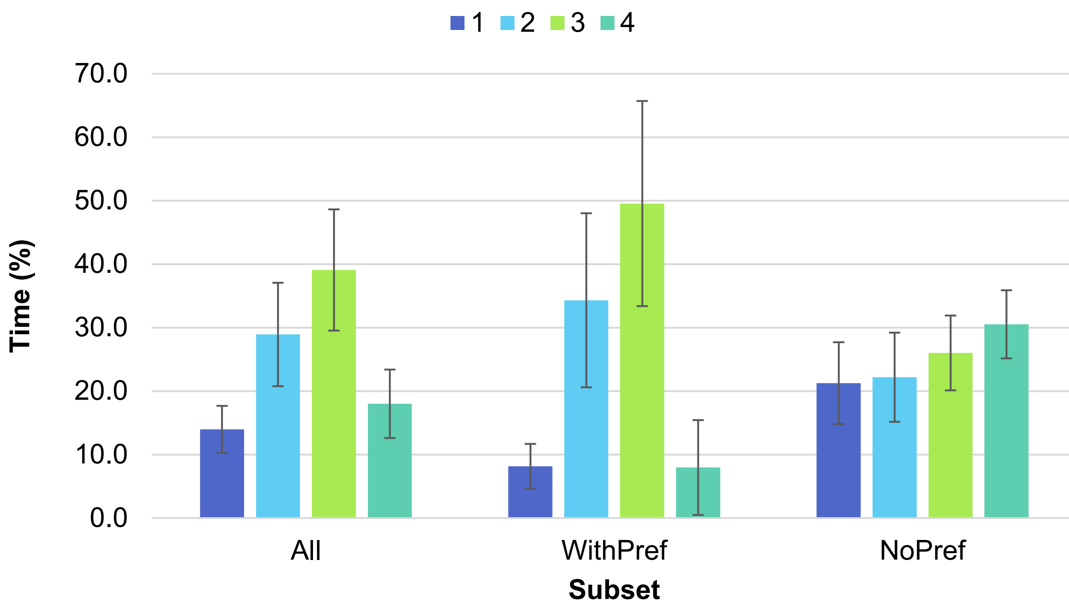


Figure 15 Estimated mean percentage of time spent in each sector for all horses (All, n = 18), those that display a reliable preference (WithPref, n = 10) and those that do not (NoPref, n = 8).
Error bars show SE. No significant differences between means were found within each subset of horses. 1 = Sector 1, 2 = Sector 2, 3 = Sector 3, 4 = Sector 4.

The results of analysis of mean percentage of time across stimulus locations for all horses (All), those that displayed a reliable preference (WithPref), and those that did not (NoPref), are presented in Figure 16. Within the All subset, there was a significant overall effect of stimulus ($p = 0.009$). During post-hoc pairwise testing, significant differences were found between the Person and Spray stimuli ($p = 0.020$), with a mean of 47.3% (SE 8.32) and 10.9% (SE 4.15) respectively. Within the WithPref subset, there was significant overall effect of stimulus ($p = 0.036$). During post-hoc pairwise testing, significant differences were found between the Person and Spray stimuli ($p = 0.014$), with a mean of 58.9% (SE 13.66) and 0.6% (SE 0.58) respectively. Within the NoPref subset, mean time did not vary significantly by stimulus ($p = 0.574$).

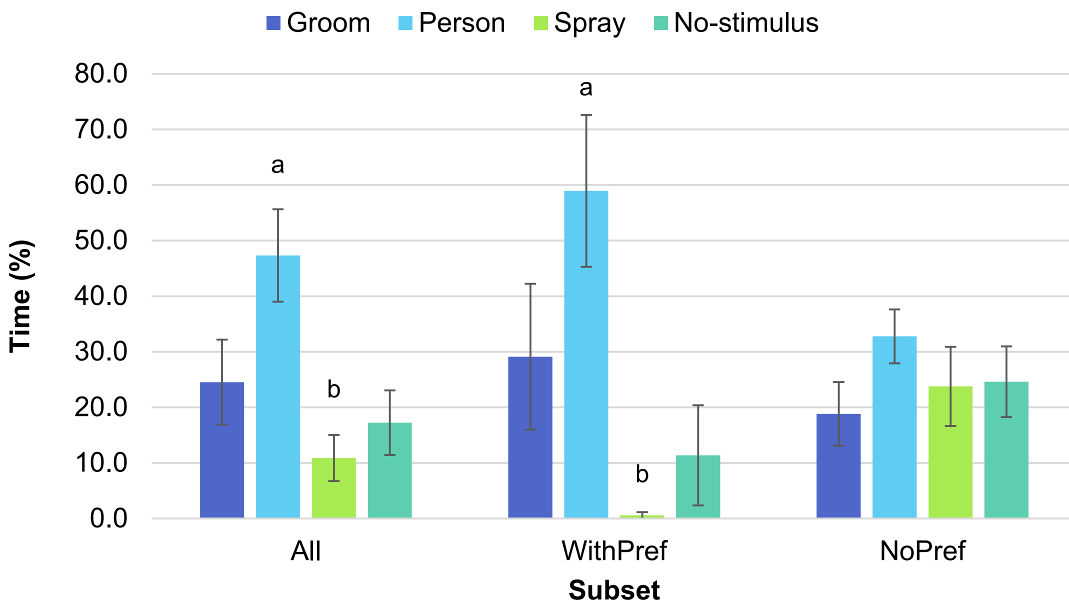


Figure 16 Estimated mean percentage of time spent in each stimulus location for all horses (All, n = 18), those that display a reliable preference (WithPref, n = 10) and those that do not (NoPref, n = 8).
Error bars show SE. For analysis of stimulus within each subset, differing letters indicate a significant difference between means. Within the All horses subset, mean time for the Person stimulus was significantly different to the Spray stimulus (p = 0.009). Within the WithPref subset, mean time for the Person stimulus was significantly different to the Spray stimulus (p = 0.036).

To further assess any impact of potential location bias on individual stimulus preferences, the mean percentage of time in each sector was normalised for habituation time for each horse (Table 14).

Table 14 Habituation-normalised mean percentage of time in each sector.

Reliable preference Y/N	Group	ID	Horse	Sector [a]			
				1	2	3	4
Y	1	1	Shane	-20.7	-29.6	**69.9**	-19.6
Y	1	3	Lucy	-1.8	-25.6	**29.1**	-1.7
Y	1	4	Spice	-3.5	-11.3	**45.8**	-31.1
Y	2	5	Semitone	-12.1	**35.6**	-22.3	-1.2
Y	2	6	Phoenix	2.9	**60.6**	-43.1	-20.3
Y	3	9	Frankie	-6.5	**84.5**	-47.6	-30.4
Y	3	12	Whitney	-2.1	-26.0	**56.8**	-28.7
Y	4	14	Pixie Fae	25.0	-50.0	-50.0	**75.0**
Y	4	16	Grace	21.0	**52.4**	-33.9	-39.5
Y	5	19	Daisy	-9.8	-17.1	**50.9**	-24.0
N	2	7	Hattie	-15.9	**35.4**	-28.5	9.0
N	2	8	Streisand	-6.6	-16.3	-10.3	**33.2**
N	3	10	Minty	-12.8	0.9	**11.2**	0.7
N	3	11	Chonty	5.1	-31.2	2.7	**23.4**
N	4	15	Toy	-11.2	**38.5**	-43.0	15.7
N	5	17	32	**33.5**	-21.7	-18.7	7.0
N	5	18	Zoe	**29.2**	-18.5	-12.7	2.0
N	5	20	Fatty	0.0	10.8	-38.1	**27.3**

Y = Yes, N = No
[a] The highest values for each row are shown in bold.

The highest sector mean time for each horse, based on the habituation-normalised sector mean (percentage of) time, was the same as that associated with the horse's most preferred stimulus in all except two horses. For those two horses (ID 8, Streisand, and ID 17, 32), they did not display a reliable preference, and the time spent during tests for two of the sectors was very close (34% versus 35%, and 51% versus 47% respectively).

5.4 Discussion

This experiment used a preference test to identify a subset of horses that displayed a preference for one of the four stimuli offered. The stimulus that was most preferred, and thus most likely to induce positive emotion, was identified for each horse for use in subsequent investigations of indicators of emotion. The spray treatment was considered the least preferred, probably an aversive stimulus, and thus most likely to induce negative emotion, based on observation of startle and escape behaviour, and due to its use as an aversive stimulus in studies on avoidance task learning (Seaman et al. 2002; Visser et al. 2003; Lansade et al. 2010). The remaining two stimuli may induce emotion with a valence that falls between that of the most and least preferred stimuli (intermediate valence).

Horses were able to make choices between stimuli-locations, but not all horses displayed a reliable preference. In this experiment, just over half of the horses displayed a reliable preference, as defined in the methods, between the stimuli-associated locations, using these methods of testing and analysis. Based on mean time across all tests, horses prefer the Person over the Spray stimulus. The same relationship, although more marked, was evident when the subset of horses that displayed a reliable preference was analysed. However, there was no preference evident in the subset of horses that did not display a reliable preference. This may be interpreted as horses for which these stimuli are salient, are likely to prefer the Person stimulus, and avoid the Spray stimulus. Regardless, combining test data across horses may be of little value when investigating preferences as they relate to indicators of emotion, and as preferences are individual.

The preferences of individual horses in the study population differed. Of those that displayed a reliable preference, the majority preferred having a person nearby, compared to wither grooming or being left alone. It may be that wither grooming by a person is overestimated as an enjoyable, relaxing stimulus for all horses (Feh et al. 1993; McBride et al. 2004; Scopa et al. 2020). The findings support the notion that what horses like cannot necessarily be assumed and may vary on an individual basis. Application of a stimulus with a presumed universal emotional response (putatively positive or negative) may not be valid. Independent verification of how an individual feels about the stimulus is required for research that is dependent on the valence of the emotion induced by the stimulus.

It is not uncommon during preference testing for animals not to make a choice, or to make unreliable choices (Fraser et al. 2018). For the horses that did not display a reliable preference, aside from chance, it may have been that the stimuli were trivial to them, that they did not like any stimulus sufficiently more than the other, or that they disliked or liked them all equally across the tests. It is also possible that they hadn’t learnt an association of the stimulus with the location and/or visual cue, were not cognitively capable of making the choices (although the cognitive abilities of horses have been largely underestimated (Hanggi 2005)), or had learnt that no stimuli were presented during test sessions therefore a choice was not necessary. It is also possible that the criteria were too conservative. Other studies have used measures of choice consistency without consideration of preference strength While the experiment achieved the aims, further research possibly using a different range of stimuli, and/or an operantly conditioned task for receiving or avoiding the stimuli would be interesting.

The test method involved a simultaneous free choice of four stimuli-associated locations, reducing the time and cost of the experiment compared to pairwise presentation in a Y- or T-maze. Refinements to the method could include presentation of the putatively aversive stimulus in an individual's preferred location during habituation, to counter any potential location bias, and more even distribution of the remaining stimuli across locations during treatment allocation. For application to other purposes, for example, estimating preferences for the population of all horses for welfare decisions, validation of the test and assessment methods against other methods, e.g., Y- or T-maze, perhaps using metrics such as number of choices and/or latency to approach, rather than time, and/or against a resource of known demand, could be considered in future. The 'location-associated simultaneous preference test' was suitable for the purpose of this study and may be an appropriate method for use in future research on emotion in animals that requires verification of the likely valence of the stimulus-induced emotion.

5.4.1 Comparison of methods of assessing overall preference

For each test, the preferred stimulus-location was determined for each horse based on where it spent the most time. Those results did not provide information about preference consistency, strength, or reliability. Based on those results, some horses would have appeared to have two or three equal preferences; others, despite spending less than half of their time in a particular stimulus-location, would have appeared to display a preference. Thus, further assessment of preference was required.

Two methods of assessing individual preference were used. The Criterion method involved using arbitrary thresholds as is common in other studies (Kirkden et al. 2006). The Choice and Time Index methods were explored to determine if they would obviate the need for use of arbitrary cut offs as in the Criterion method. The Choice Index was based on the relative number of times that each stimulus-location was preferred or chosen in all tests and disregards the time values. The ‘most preferred stimulus’ was found to be the same using the Choice Index for horses that displayed a consistent choice. For the other horses, the Choice Index identified that they did not display a preference (Choice Indices ≤ 0), which also agreed with the preference consistency results using the Criterion method. Thus, the Choice Index method may combine determination and interpretation of the ‘most preferred stimulus’ and the ‘Preference consistency’ into one, without the use of an arbitrary cut off value. However, it is unclear if the magnitude of the Choice Index was meaningful in assessing preference strength.

Minty was identified as having a weak preference for the No-stimulus treatment using the Choice Index and Criterion methods, but the preference strength did not reach the criterion for preference reliability. It may be that a cut off is also required for the Choice Index magnitude (for example an arbitrary > 50%) to interpret preference strength and reliability, which would then negate the advantage over using an arbitrary method such as the Criterion method. Additionally, the Spray stimulus was not preferred by any horses based on Choice Index, despite some horses spending much of their time in the Spray location in one or two tests. The value of the Choice Index versus the Criterion method in evaluation of the putatively negative stimulus is not certain.

The Time Index incorporated the relative time spent in each stimulus-location for each horse averaged across all tests, but disregarded the preferences shown in each test. The results of the Time Index differed slightly from assessment of preference using the Criterion method. Minty was not shown to have displayed a preference using the Time Index method (Time Indices ≤ 0), which supports the suggestion that, although Minty chose consistently, he did not display a strong preference. Streisand was found, with the Time Index, to have a very weak preference, but was not found to have a preference using either the Choice Index or Criterion methods.

The Criterion method used in this study gave a more conservative determination of preference than either Index calculation. For the purpose of identifying horses that displayed a reliable preference (i.e., consistently preferred on most occasions), the Criterion or Choice Index methods of preference assessment may be suitable. However, the utility of the Time Index method in assessing the strength of preference warrants further consideration.

5.4.2 Influence of location bias on apparent stimulus preferences

In the absence of any preference or confounding factors, it would be expected that time would be evenly distributed between sectors. However, a location bias towards Sector 3 was apparent during habituation. Habituation data across all horses showed that, on average, horses spent significantly more time in Sector 3 during habituation sessions (Figure 14). This location bias may be because Sectors 1 and 4 had a more acute angle of entry than Sectors 2 and 3, although this was somewhat mitigated by having sector ‘o’. Sectors 1 and 4 also bordered the outside of the arena along an edge which may be a more vulnerable location, especially for a prey species. Interestingly, there was no such bias during habituation for Sector 2.

On an individual basis, no horses would have met the criteria for displaying a reliable preference for location during habituation. Three horses displayed a moderately strong, consistent preference, which was for Sector 3. However, none of those horses went on to display a reliable preference during testing.

If a location bias had persisted through training and testing, then a similar result would be expected in the test data. However, during preference testing, there was no significant difference between the mean time spent in each sector for all horses, or those that did or did not display a reliable preference (Figure 15). Location time had changed between habituation and preference testing, presumably due to the association with stimuli during training.

When compared across stimuli, there were significant differences in the time spent in each stimulus-location for those horses that displayed a reliable preference, but not for those horses that did not display a reliable preference (Figure 16). Overall, horses spent significantly more time in the Person stimulus-location, than in the Spray stimulus-location. That same relationship was not evident in the subset of horses that did not display a reliable preference as defined here. Their time was more evenly spread across the stimuli and sectors, as expected if they either had no preference for one of the stimuli, or if there were no other confounding factors affecting location. This provides further assurance that the location bias that was apparent during habituation was overcome by the training and no other confounding factors were affecting location.

Despite the lack of evidence of a location bias during testing, visual examination of the test data raised suspicion of a location bias for some horses. For example, Group 1 horses preferred Sector 3 100% of the time in all tests, and for horses that displayed a reliable preference, the most commonly preferred location was Sector 3 (5/10); the Person stimulus, which was the most commonly preferred stimulus (6/10), was presented most often in Sector 2 (4/10), followed by Sector 3 (3/10). Any potential individual location bias was then accounted for by normalising the test data for the time spent in each sector during habituation on an individual level. Based on the habituation-normalised data (Table 14), the 'most preferred stimulus' did not change for horses that displayed a reliable preference.

Horses display difficulty in reversal learning, which may confound results of preference testing during attempts to control for a side or position bias by using a cross-over design (Grandin et al. 1994; Sappington et al. 1997). In the current experiment, although the position in which each stimulus was presented was randomised over all horses, the location of each stimulus was fixed for each horse to maximise spatial association and eliminate reversal learning problems. Although attempts were made to control systemic confounds, such as social proximity and food/grass cover, the impact of individual side biases (laterality), or other factors that may consistently impact location preference, were not ameliorated.

It is acknowledged that overall, there was an unequal allocation of stimuli to sectors that resulted from using a modified Latin square method of random allocation, and that the subsequently determined preferred location based on mean time during habituation for all horses (Sector 3) coincided with the highest allocation of the No-stimulus and Person (which was also the most commonly preferred stimulus) conditions. Two possible refinements of the study design would be to present the putatively aversive stimulus (Spray) in the preferred location for each horse (as determined by highest mean time during habituation) to counter any apparent individual location bias in favour of the other stimuli, as has been applied previously with Y maze preference testing (Arnold 2005), and to limit the maximum number of times that a stimulus may be allocated to a sector during random allocation to ensure a more even distribution.

5.4.3 Limitations

Nature of stimului

This research did not set out to determine what horses do or do not enjoy. It was designed to identify a subset of horses that cared about the offered stimuli for use in the subsequent experiment. The treatments differed in many aspects, including the proximity and/or involvement of a human, and degree of control the horse had over the treatment. For example, the no-stimulus and spray treatments did not involve a human; interaction with the human during the Person treatment was voluntary, whereas the groom treatment was initiated by the human and not voluntary, as is the case with other studies (Lansade et al. 2018). Horses' individual preferences for the treatments may have been related to these aspects. Further research on characterisation of horse preferences for putatively positive stimuli in general would be interesting.

Social contact, access to roughage, and free movement are basic needs of horses (Krueger et al. 2021). Aligning also with Maslow's hierarchy of needs (Yeates et al. 2008), it may be that for horses in impoverished situations, their utilisation and preferences for enrichment experiences may be altered (Jørgensen et al. 2011). The subjects involved in this study were not considered to be impoverished, and stimuli were chosen carefully so that negative states were not induced unintentionally. For example, the use of social contact as a positive stimulus, which has been used in studies in horses (Briefer et al. 2015b), requires the prior imposition of social restriction or isolation, and horses with stereotypies, which may be associated with impoverished environments, may respond paradoxically (Normando et al. 2007). Care is advised when assessing preferences in animals where basic needs are not met or are deliberately manipulated (Fraser et al. 2018).

Study design

Design and interpretation of preference testing can be complex and relate to the specific research questions of interest (Kirkden et al. 2006). The current study may be classified as a non-exclusive, between-motivations choice test, which is suitable for assessment of preferences but may be less suited to assessment of strength of preference between the stimuli (Kirkden et al. 2006). This was recognised as an inability to rank the motivation for the intermediate stimuli (i.e., the two stimuli that were not the most or least preferred). Inclusion of measures of effort to obtain/avoid a resource/stimulus e.g., operant tasks and elasticity of price/income demand, could be considered to further explore the strength of preference between stimuli, but would also add to the study complexity, time involved, and require a larger pool of subjects (Fraser et al. 2018; Smith et al. 2018).

The order of stimulus delivery was the same for all horses during training sessions (PGSC). This was done, in part, to mitigate potential overshadowing effects of the Spray stimulus on the other stimuli, and so that the Spray stimulus was not the last one encountered in the arena, to reduce negative associations with the arena. Similarly, horses spent at least 10 minutes in the holding pen between treatment exposures to limit potential for carry over effects. It seems unlikely that an association between treatments would have formed over such a long period of time (McGreevy 2012), and preferences for the No-stimulus and Groom stimuli were apparent in some of the horses that displayed a reliable preference. Thus, the impacts of Spray on other treatments were considered negligible.

No stimuli were presented during test sessions, which meant a heavy reliance on the association of the visual and spatial cues with the stimuli during training, although the association was reaffirmed by conducting a single training session prior to each test. One limitation of the design of this choice test was that it was largely unknown if horses had acquired an association of the stimulus with the visual and spatial cues, or how long that association lasted. Although, horses have been shown to discriminate symbols during preference testing (Mejdell et al. 2016) and the distribution of horses' time was not equal across the stimuli-locations during tests, contrary to expectation if an association had not been formed. It is concluded that some learning had occurred.

The number of tests was reduced for four horses due to inclement weather. For three of those horses, the outcome would not have changed if a fourth test had been included. For the remaining horse, Minty, a fourth test would not have changed its preference, but may have led to improved preference strength. Thus, the impact of a reduction in tests for three horses was conservative for one and none for three horses. It may be possible to reduce the number of tests from four to three in future studies that rely on preference consistency.

Influences on preference

Lansade and Simon (2010) suggested that learning in horses is highly task dependent and impacted by tactile sensitivity and fearfulness, amongst other temperament traits that are stable over time (LeScolan et al. 1997; Lansade et al. 2008c). It could be that in the current study differences in presence or absence of preference, and/or preferred stimulus reflect differences in learning ability and temperament between horses. Examination of the habituation data (Figure 14) compared to the test data (Figure 15) showed that horses' behaviour had changed from the habituation phase to the test phase. Learning had occurred in horses and preferences had developed in some. This may be a reason that not all horses displayed a preference, and reinforces the idea that preferences are likely to be unique to the individual.

Horses' preferences may be influenced by physiological state, environment, timing (e.g., the time of day for peak Grooming demand may not coincide with the time for peak Person demand), and initial familiarity with the stimulus (Kirkden et al. 2006; Fraser et al. 2018). Although within-subject factors such stage of oestrous were not specifically controlled, obvious signs of oestrous were not noted during testing. Testing was conducted over several days with as many aspects of the environment controlled as possible in an outdoor setting. Due to the conservative criteria used to assess preference, it is considered unlikely that day and/or environmental conditions would have led to horses displaying a preference where none existed. Peak times of demand for the offered stimuli were unknown. Any issues with initial familiarity, either positive or negative, were addressed via use of habituation, previously unfamiliar people, and training. Repetition of the study in a climate-controlled environment, with horses outside of the breeding season could be considered. Horses' preferences have been shown to be stable a year later (Lansade et al. 2018), which lends confidence to the findings of the current study.

The stimuli used in this experiment were classified as non-substitute alternatives, that is, they may not have the same motivational basis. Consideration of the metric used to determine preference should be given because satiety for each alternative may be achieved in different ways e.g., amount consumed, time, number of engagements, and/or at different rates (Kirkden et al. 2006). It is not known how satiety is achieved for the Person, Grooming, and No-stimulus conditions, but time may be a logical measure and it has been used in previous studies

(Janczarek et al. 2018b; Scopa et al. 2020). Naturally occurring bouts of allogrooming in horses occur for 1-5 minutes every few hours (Feh et al. 1993; van Dierendonck et al. 2004; Wolter et al. 2018; Shimada et al. 2020). The stimuli used in the current study were not available to the horses outside of the study. The duration used in the study was not expected to reach satiety for the non-aversive treatments, and thus was not expected to impact the results. Additionally, habituation to the Spray stimulus was not observed behaviourally or in the test data for horses that displayed a preference.

Examination of the potential effects of chance on the outcomes is warranted. While roughly equal numbers of horses displayed a preference or no-preference, the chances of a horse falsely displaying a preference were much less than a horse falsely displaying no-preference. The number of horses displaying no-preference may be more likely to be inflated by chance. The inclusion of four tests rather than three reduced the number of errors in each group. The distribution of the five preference outcomes (Groom, Person, Spray, No-stimulus, No-preference) was not due to chance.

A premise of this study, and dependent studies, links emotional valence induced by a stimulus to an animal's relative preference for the stimulus. It could be argued that if a horse did not find any of the stimuli pleasant, then its preference may actually be for the least aversive stimulus (aversion test). It is acknowledged that, in the case of an aversion test, the underlying emotion may not be an absolute positive, but more positive than that experienced during the other stimuli (and vice versa for negative emotion). It is also acknowledged that the preferences, as determined by this experiment, were relative preferences among what was offered.

5.4.4 Conclusion

In a novel design of a free choice preference test, just over half of the horses indicated their preferred stimulus by choosing to spend more time in an area associated with that stimulus. Two methods of assessing preference were described in this experiment. Little advantage was found with the Index methods over the Criterion method. Of horses that displayed a reliable preference, the Person stimulus was most preferred, over the Groom and No-stimulus conditions, and the Spray stimulus was least preferred. However, preferences varied on an individual basis.

Future research on indicators of emotion in animals should not assume that exposure to a given stimulus will result in a universal response and induce emotion of the presumed valence. Studies should verify how an individual perceives a given stimulus prior to measuring potential indicators of emotion. Although the limitations of this experiment are acknowledged, the ‘location-associated simultaneous preference test’ described here is one solution which could be applied to horses, and potentially other species, in future research.

Chapter 6 Investigation of indicators of emotion in horses

6.1 Introduction

In the final experiment, the horses that were identified in experiment two (Chapter 5) as displaying a preference among four stimuli using the 'location-associated simultaneous preference test' were used to investigate physiological, behavioural, and facial responses as indicators of emotion. This chapter briefly introduces the experiment and presents the general methods used. The more specific methods, results, and discussion relating directly to each parameter category are presented in the following three chapters: physiological measures (Chapter 7), gross body behaviours (Chapter 8), and head and facial movements (Chapter 9). In Chapter 10 the results from Chapters 7 to 9 are brought together and interpreted with respect to the arousal and valence dimensions of the subjective component of emotion. Discussion and critique of the findings and research approach are also presented in Chapter 10. Multivariate analysis of the results of the final experiment are presented in Chapter 11.

For this experiment, several assumptions were accepted. A preferred stimulus is more desirable than the other stimuli and induces a more positive emotion (Humphrey 1972; Boissy et al. 2007b; Dawkins 2008). The Spray stimulus is likely to be aversive and induce negative emotion, as it was never preferred by any horse that displayed a preference, and has been used as an aversive stimulus in studies on avoidance task learning (Visser et al. 2003; Lansade et al. 2010). The other two stimuli are likely to induce emotion with a valence that falls between that of the most and least preferred stimuli (intermediate valence).

The aim of this study was to investigate how the physiological and behavioural responses of horses change whilst undergoing a pleasant experience, compared to a mildly aversive, and an intermediate experience. It was hypothesised that the preferred stimulus would increase parasympathetic nervous system activity, assessed using eye surface temperature and heart rate variability (Boissy et al. 2007b), and induce behaviours associated with a relaxed state.

6.2 Materials and methods

All procedures were conducted in accordance with the Massey University Code of Ethics for Animals (MUAEC approval 10/116).

6.2.1 Animals and housing

Eleven horses from a teaching herd maintained at the Massey University Veterinary Large Animal Teaching Unit, Manawatu, New Zealand, that were found in the previous experiment to display a preference for one of the stimuli, were used for this experiment (Table 15). One horse (ID 10 Minty) that displayed a consistent, but weaker (strength 46.4%) preference in the previous experiment was included to have more subjects for statistical power, although he did not meet both of the reliable preference criteria.

Throughout the experiment, horses were kept at pasture in established management groups so that social bonds were not disrupted. Horses were accustomed to being handled. Their diet consisted of grass and hay, with water provided *ad libitum*. Ages ranged from 5-22 years old. There were nine mares, one gelding, and one ovariectomised mare; seven were Standardbred and four were Thoroughbred or Thoroughbred cross breeds. Six horses had a preference for the Person, two for the Groom, and two for the No-stimulus conditions.

Care was taken when handling horses to avoid horses inadvertently forming negative associations with the environment or personnel involved. No horses were excluded from the study, as none were affected by significant illness or injury, or displayed abnormal behaviour.

Table 15 Details of the horses used in experiment three, preference and management group.

ID	Name	Sex	Age (years)	Breed	Colour	Most preferred stimulus	Group
1	Shane	G	6	SB	Brown	G	1
3	Lucy	M	10	SB	Bay	P	1
4	Spice	M	5	SB	Bay	P	1
5	Semitone	M	22	SB	Bay	G	2
6	Phoenix	M	11	TB	Bay	P	1
9	Frankie	M	5	SB	Black	P	2
10	Minty	M	14	SB	Bay	C	1
12	Whitney	M	17	TB	Chestnut	P	2
14	Pixie Fae	M	13	SB	Bay	G	2
16	Grace	M*	17	TBx	Chestnut	P	1
19	Daisy	M	15	TB	Bay	C	2

ID = Identification number, G = Gelding, M = Mare, * = Ovariectomised, SB = Standardbred, TB = Thoroughbred, x = Cross, P = Person, G = Groom, C = No-stimulus

6.2.2 Facility setup

The experiment was conducted at the indoor teaching and research facility on the Massey University Veterinary Large Animal Teaching Unit farm where the horses were kept, and in the set up as shown in Figure 17. The experimental horse was loosely restrained in the stocks by a rope in front and behind to limit ambulation, and to enable behavioural and physiological measurements to be taken with limited effect of exercise. The building had a concrete floor and lighting was by natural and artificial means. One infrared thermography (IRT) camera (ThermaCam S60, FLIR Systems AB, Danderyd, Sweden) was positioned on the right side of the horse to record the temperature of the eye area. Four video cameras were positioned to record body (Sony DCR-SR85E, Auckland, New Zealand; side, front, and oblique head views), head, and facial movement, behaviour (Canon Legria HFM31, Sydney, Australia; high-definition left side of head view), and vocalisations. The camera views of a horse in the experimental setting are shown in Figure 18, Figure 19, Figure 20, and Figure 21.

Figure 17 Diagram of facility setup.

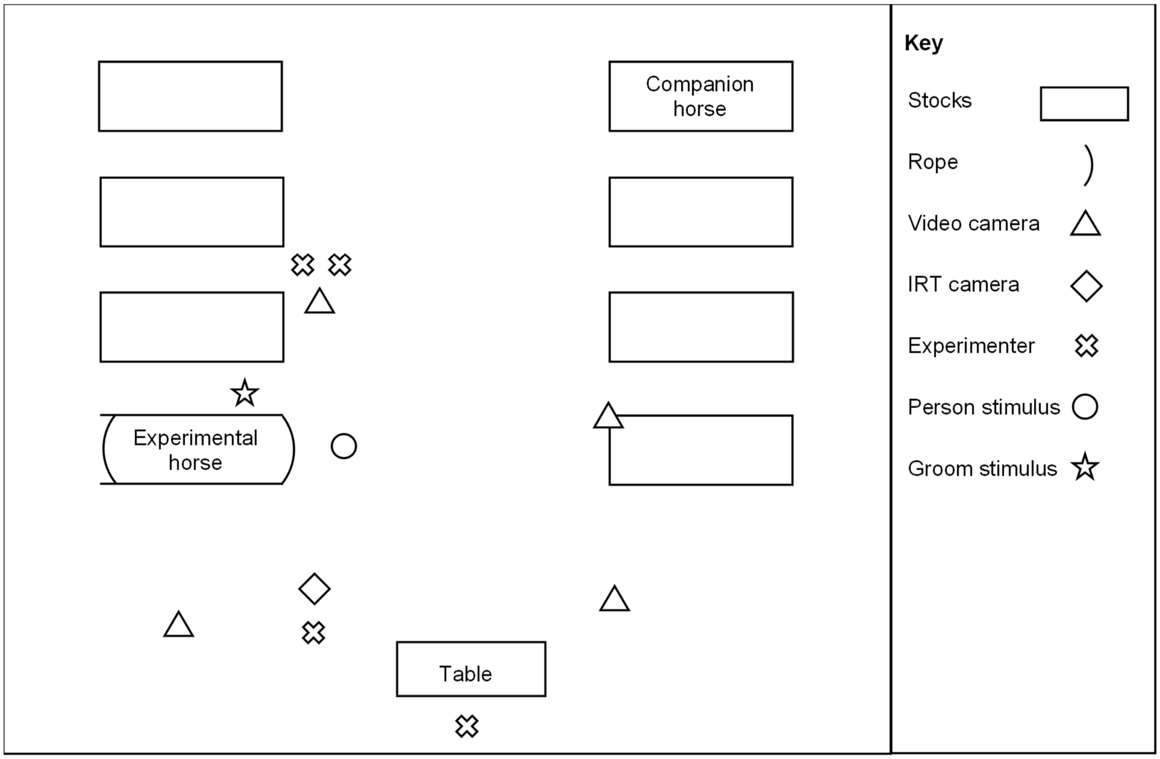


Figure 18 View of the horse as recorded by the side camera with a fisheye lens.

Figure 19   View of the horse as recorded by the front camera.

Figure 20   View of the horse as recorded by the oblique camera.

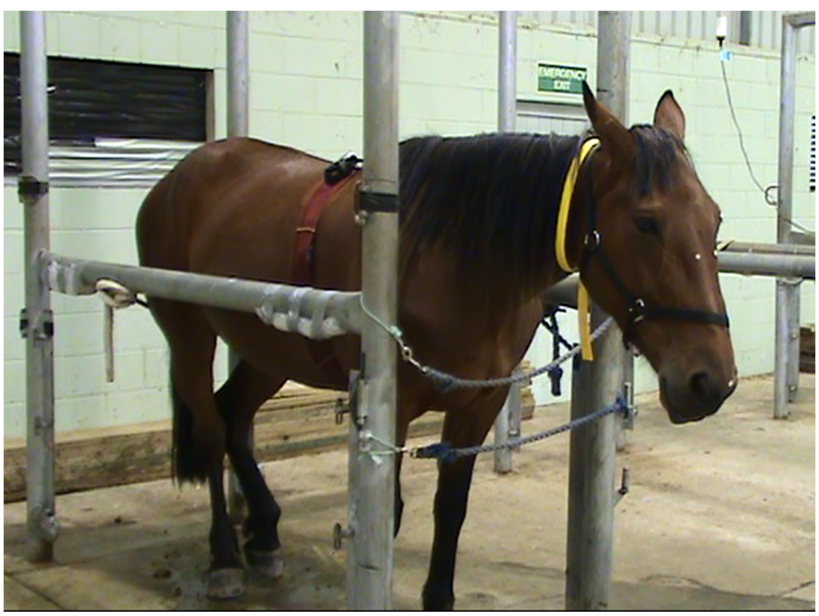
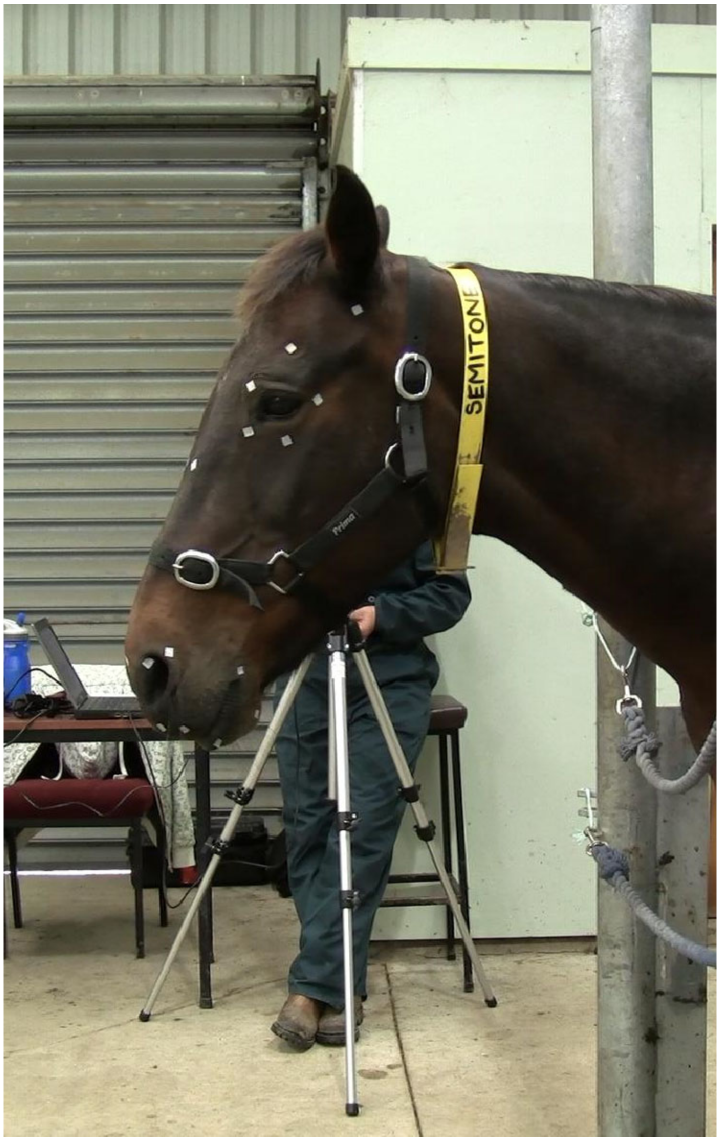


Figure 21   View of the horse as recorded by the head camera.



6.2.3   Treatments

The Person, Groom, Spray, and No-stimulus treatments were the same as the stimuli used during the previous experiment, without the visual cue (Table 16). A different person administered the Groom and Person treatments, and was consistent for each horse. For the Spray treatment, a single spray was delivered at 20 second intervals for a total of three sprays, i.e., a spray was applied at time 0 s, 20 s, 40 s from treatment start. A period of 140 seconds followed the third spray to make a total of three minutes for the treatment. The spray was not repeated continuously during the three-minute treatment period to minimise potential for habituation to it, and to reduce the potential for harm to the horse if attempts were made to escape.

During initial study design, consideration was given to inclusion of a control situation, for example the No-stimulus treatment, for which the effects observed across treatments and horses could be compared. An ideal control treatment would induce a neutral emotional valence, that is, it would not elicit a positive or a negative emotion and would be regarded the same by all subjects. However, a control or no-stimulus situation may not elicit the same emotional valence or arousal across all animals, because animals' emotional responses are specific to the individual animal. There are no gold standard treatments for this type of research with which comparison can be made. Thus, a universal control situation was not possible to achieve, and the No-stimulus treatment cannot be thought of as one. Consequently, the No-stimulus condition was included as a treatment rather than a control. It was the preferred treatment for some of the horses, which highlights the point that it is not regarded equally by horses

Table 16: Name and description of treatments.

Treatment	Description
Person (P)	A previously unfamiliar person stood motionless 1m in front of the horse, with a relaxed posture, not interacting or making eye contact with the horse. The horse could choose to interact with the person and was able to sniff or lick them (horse-initiated contact). The person stepped back out of reach if the horse tried to bite. The same person was used for each repetition for an individual horse.
Groom (G)	A previously unfamiliar person (different to P) approached the horse and groomed the left wither area using a rubber grooming tool in a circular motion (human-initiated contact). The same person was used for each repetition for an individual horse.
Spray (S)	A single lightly scented aerosol spray was emitted every twenty seconds for three sprays (if deemed safe to do so) from a remote-controlled spray unit (Petsafe Spray Remote Trainer System SRT-200, Radio Systems Corporation, Knoxville, USA) located behind the right elbow. The spray device was placed under a surcingle on the right side of the horse's chest, midway between ventrum and dorsum and was worn during all sessions.
No-stimulus (C)	Horses were standing in stocks. No additional stimulus was present.

6.2.4 Experimental procedure

Following habituation and training, a horse's preferred stimulus (as determined in Chapter 5, and given in Table 10) was used to induce positive emotion, the spray stimulus was used to induce negative emotion, and the other two stimuli were used to induce emotion with a valence in between (intermediate), while behavioural and physiological variables were measured.

Horses were habituated to the facility and equipment and tested in two groups (Table 15). At the start of each day, a group of horses were led to the experimental facility and put into the covered holding pens with water and hay available. While in the holding pen, a heart rate monitor (Polar Equine RS800CX, Polar Electro Oy, Helsinki, Finland) and a spray device (Petsafe Spray Remote Trainer System SRT-200, Radio Systems Corporation, Knoxville, USA) were placed under an elastic surcingle, with the spray device on the right side of the horse's chest, midway between ventrum and dorsum. Fifteen small white markers (1 cm x 1 cm flexible, breathable self-adhesive dressing; Fixomull, Smith and Nephew, Sydney, Australia) were placed on prominent anatomical features of the head to facilitate close observation of changes in facial features (Figure 21 and Figure 24). They did not seem to disturb the horses. Horses were bought into the stocks area in pairs to prevent social isolation. The experimental horse's head was not restrained whilst in the stocks to allow free movement of the head and neck. The horses were led back to their paddock at the conclusion of each day's events.

Although the horses were familiar with the facilities, the horses were given three 30-minute sessions to ensure habituation to the experimental environment and equipment. The horses were trained to back into the stocks using positive and negative reinforcement. Horses had access to hay during habituation sessions, but not during the experiment sessions. Horses that displayed anxious or fearful behaviour (vocalising, head high, tight lips, taut neck muscles, increased reactivity to sounds or movement, heart rate >60bpm) could be given further habituation sessions, however, no horses required this.

Each experiment session consisted of repetitions of the four different treatments for three minutes' duration, and had a total duration of approximately 50 minutes. The session was repeated three times for each horse, with at least 24 hours between sessions. The schedule of events for each session is given in Table 17 and encompassed 22 periods of time. The

Spray treatment was applied once per session (over periods 9-11), while the Person and Groom treatments were applied two times, once before and once after the Spray treatment. The No-stimulus treatment was applied seven times, three times before and four times after the Spray treatment, as it was interspersed between the other treatments. To account for the possibility of a negative carryover effect from the Spray treatment (overshadowing), the other treatments were applied before and after the Spray treatment, and the treatment order was partially fixed. The order in which the Person and Groom treatments were given was randomised for each horse. Food was given at the end of each session to counteract (counter-condition) any negative association with the experimental area that may have been formed due to the aversive stimulus, and to facilitate easy return of the horse to the stocks in subsequent sessions. A bucket containing a measured amount of concentrated feed was offered to the horse. The bucket was placed on the ground in front of the horse where it was accessible to the horse.

Table 17: Experiment schedule of events for each 50-minute session for each horse.

Period	Treatment	Duration (seconds)
1	Pre-treatment – 10-minute settling period	600
2	No-stimulus	180
3	Approach to deliver next treatment	2-7
4	Person or Groom – randomly allocated	180
5	No-stimulus	180
6	Approach to deliver next treatment	2-7
7	Person or Groom – alternate from period 4	180
8	No-stimulus	180
9	Spray 1	20
10	Spray 2	20
11	Spray 3	140
12	No-stimulus	180
13	Approach to deliver next treatment	2-7
14	Person or Groom – randomly allocated	180
15	No-stimulus	180
16	Approach to deliver next treatment	2-7
17	Person or Groom – alternate from period 14	180
18	No-stimulus	180
19	Approach to deliver food bowl	8-37
20	Food	180

Period	Treatment	Duration (seconds)
21	Remove food bowl	10-18
22	No-stimulus	180

6.2.5 Measures

Heart rate (HR), heart rate variability (HRV), eye temperature (ET) and behaviour (video) were recorded continuously throughout each session. The method for each variable is given under the relevant section in Chapters 7, 8, and 9. In total, seven physiological parameters, 21 body movement, and 34 facial movement behaviours were measured.

6.2.6 Overview data handling and analysis

The data set was comprised of results from 11 horses with three replicates per horse. The results from the two Groom, two Person, two of the No-stimulus, and the Spray treatments for each replicate were analysed. A No-stimulus treatment was randomly selected from the three periods occurring before the Spray treatment and another one was randomly selected from the first three periods occurring after the Spray treatment. An ordinal variable was created to code the individual preference for each treatment as Preferred, Intermediate, or Aversive, based on the results of preference testing described in Chapter 5 of this thesis.

Continuous data were described by median, inter-quartile range (IQR) and range, and are presented as boxplots in Appendix A.4, A.5, and A.6. Categorical data were described by the number and percentage of observations, and are presented in Appendix A.4, A.5, and A.6. No further statistical tests were conducted during the exploratory data analysis phase because the data violated assumptions of independence. Rather, mixed-effects models were used to examine the relationship between preference, or treatment, and each variable (Zuur et al. 2009). These models were necessary to account for the impact of preference or treatment on each of the outcomes, while accounting for the repeat measurements on horses on different days, and correlation between successive measurements (i.e., temporal association between HRV measures), giving conservative estimates of the standard errors (SE). As the model accounts for repeated measures no further correction to SE or p value was necessary.

When developing the models, it was not possible to consider the impact of both preference and treatment in the same model, because the Spray treatment was always the least preferred, which created collinearity and, as a result, the parameters could not be estimated. Consequently, the impact of preference and treatment on each variable was separately examined. Two random effects (horse and replicate within horse) were added to the models to account for the baseline in each session, as there was some evidence of a day effect. Results from the Preferred category were compared to those from the Intermediate category and to those from the Aversive category. Results from the No-stimulus treatment were compared to those from the Person, Groom, and Spray treatments.

In the final models, the assumptions of normality and homoscedasticity of the errors were assessed by graphical analysis of the residuals. Where required, data transformations were performed, and models re-run using the transformed variable as the outcome.

Statistical analysis was performed in R: A Language and Environment for Statistical Computing (version 4.1.1, R Core Team, The R Foundation for Statistical Computing, Vienna, Austria). The significance of relationships was assessed using the log likelihood ratio test and Wald test statistic. Unlike Chapter 4 and Chapter 5, where Bonferroni correction was applied following post-hoc pairwise testing of ANOVA results, no post-hoc correction of p values was applied following analysis using mixed effects models. Instead, a trade off approach was taken when considering statistical significance in conservative attempt to reduce type II versus type I errors, and sacrificial loss of power. Statistical significance was set at $p \leq 0.05$, but due to subtleties of changes between more positive treatments, in some cases, the data for $p \leq 0.10$ are also discussed as showing tendency towards significance. This approach has been used previously in behavioural and ecological research for example, Lansade et al. (2018). All mean values given are based on model estimates, except where stated. In the case of log transformed variables, these are reported on the original scale.

Chapter 7 Physiological indicators of emotion in horses

7.1 Introduction

This chapter investigates selected physiological responses as indicators of emotion in horses. Emotional arousal or excitement, whether due to positive or negative emotion, may be indicated by increased activity of the sympathetic branch of the ANS. Positive emotion may be indicated by increased activity of the parasympathetic branch (Stewart et al. 2011; Zebunke et al. 2011). The assumption is made that PNS predomination in animals is positive, as it has been linked to social and psychological well-being in humans (von Borell et al. 2007; Kok et al. 2010).

Briefly, the physiological measures that were investigated in this study, heart rate (HR), heart rate variability (HRV), and eye temperature (ET), were selected because they are non-invasive indicators of the ANS (von Borell et al. 2007). Time domain indices of HRV, including mean IBI, mean HR, SDNN, RMSSD and RMSSD:SDNN ratio, were selected for analysis in this experiment due to the duration of measurement (nature and length of treatments) and anticipated variation in respiratory rates. HRV is discussed in more detail in Chapter 3.

Eye temperature, measured using infra-red thermography (IRT), has been validated in cattle and sheep as a non-invasive indicator of pain, fear, and stress (Stewart et al. 2008b; Stewart et al. 2008a; Stewart et al. 2010b). An acute drop in eye temperature is thought to be caused by vasoconstriction due to a predominance of the SNS with the negative emotional experiences (Stewart et al. 2008a). It is theorised that ET might increase in response to positive emotion (Church et al. 2009), possibly through stimulation of the PNS.

Eye temperature in horses has been investigated in relation to ill health, fitness and performance, stress/fear due to competitions, equipment, management, and housing (Jansson et al. 2021). Most of the studies have involved situations of likely high arousal, negatively valenced emotion.

The aim was to determine if HR, HRV, or ET varied with preference for a treatment (Preferred, Intermediate, Aversive), or varied with treatment (Groom, Person, No-stimulus, Spray). It was hypothesised that a predominance of the PNS with positive emotion may lead to vasodilation and an increase in eye temperature, and an increase in RMSSD or RMSSD:SDNN ratio(Church et al. 2009). It was hypothesised that a predominance of the SNS with negative emotion may lead to vasoconstriction and a decrease in eye temperature, an increase in HR, and a decrease in SDNN or RMSSD:SDNN ratio.

7.2 Materials and methods

The interval between heart beats (IBI) was measured using Polar Equine heart rate monitors (RS800CX, Polar Electro Oy, Helsinki, Finland) fitted to the horse using an elastic surcingle placed around the chest and with conductive gel on the sensors.

Measurement of IBI can be prone to errors due to disruptions with the electrode contact, muscle fasciculations, T waves mistaken for R waves, or second degree arterio-ventricular heart block (normal in many horses at rest), which produce artefacts in the HRV parameters (von Borell et al. 2007). Therefore, the data set was corrected. Rather than using the proprietary software's (Polar Pro Trainer 5 Equine Edition version 5.40.173, Polar Electro Oy, Helsinki, Finland) undisclosed method of error correction, the IBI data points with values ≥ 5750ms (equivalent to HR of ≤ 10bpm) and ≤ 200ms (HR ≥ 300bpm) were removed, as these are physiologically unlikely in normal horses (Hamlin et al. 1972; Rezakhani et al. 2005). Thirty-nine data points were replaced with missing values on this basis. In no treatment period did this equate to a 5% or more error rate, and neither were there any occasions where there were three or more consecutive errors in a row, thus all data sets were able to be included in the analysis, as per recommendations by von Borrell et al (2007).

The mean IBI, mean HR, SDNN, RMSSD and RMSSD:SDNN ratio were calculated from the error corrected data, over 10 second periods, using Microsoft Excel 2010 (Zebunke et al. 2011). A 10 second period was chosen as a compromise between 60Hz which is the frequency which the data was recorded and a longer period (e.g., 60 seconds) to maintain inherent variability within the data.

Thermal images of the right eye were recorded continuously using an IRT video camera (ThermaCam S60, FLIR Systems AB, Danderyd, Sweden) positioned on the right side of the horse, at a 90° angle to the sagittal plane and at a distance of two metres. The resolution of the IRT camera is 320 x 240 pixels (76,000 picture elements), a thermic sensitivity of sensitivity < 0.06°C, up to 60 Hz image frequency, and an accuracy ±2°C or 2% of reading. Ambient temperature and humidity were recorded at the start of each session to correct for atmospheric conditions. The maximum temperature at the medial canthus/lacrimal caruncle of the eye and average temperature of the area (Figure 22) was calculated two to three times per minute using image analysis software (ThermaCAM Researcher Pro version 2.10, FLIR Systems AB, Danderyd, Sweden). Clear images that contained a minimum of five pixels were used.

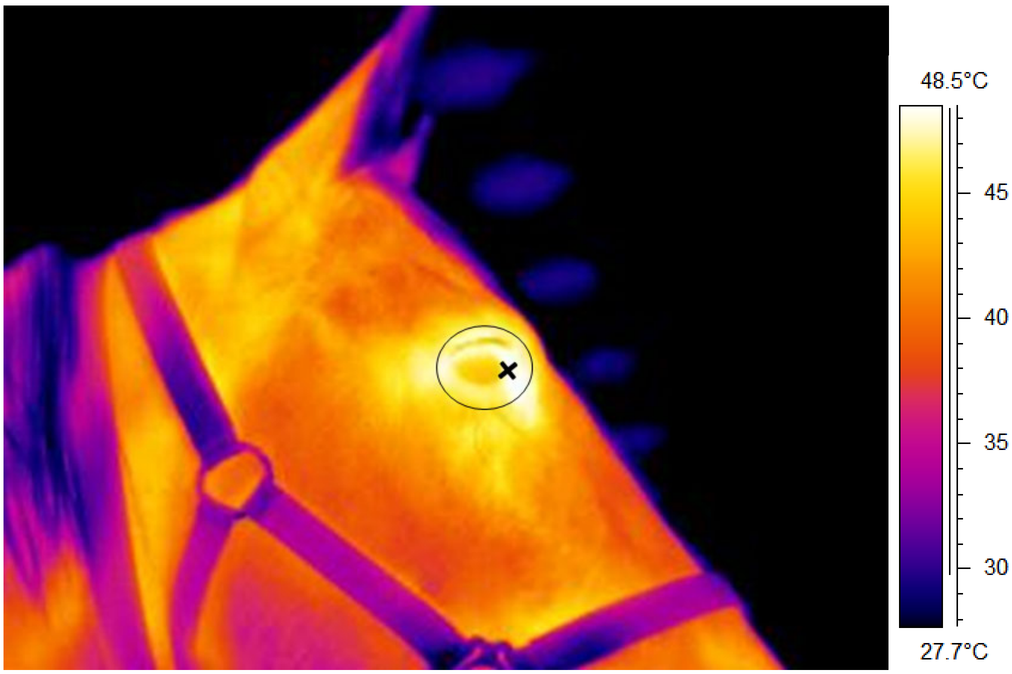


Figure 22 Thermal image of a horse's head showing the position at the medial canthus of the eye (lacrimal caruncle) where maximum eye temperature was taken (x) and the area over which the average eye temperature was calculated (oval).
ºC = degrees Celsius.

7.2.1 Data Analysis

The relationship between five heart rate (HRav) and heart rate variability (IBIav, SDNN, RMSSD and SDNN:RMSSD ratio) parameters and treatment or preference was examined using a linear mixed effects model. The model also included an auto-regressive covariance structure to allow for the fact that data was calculated every 10 seconds, thus it was likely to be correlated, and was also adjusted for repeated measures. The final models use the natural log transformation of HRav, SDNN, and RMSSD. All figures from transformed variables are the back transformed log(n) predicted from the model.

The impact of preference and treatment on average and maximum eye temperature was examined separately using a linear mixed effects model. Humidity and air temperature varied within and between days and, therefore, the models were extended to examine the impact of these variables on each of the eye temperature outcomes. Humidity was subsequently removed from the model as it did not have a significant impact.

7.3 Results

7.3.1 Heart rate and heart rate variability

The observed data for HR, IBI, SDNN, RMSSD, and RMSSD:SDNN ratio are presented in line graphs and box plots in Appendix A.4.1. The statistically significant results of the analysis of HR and HRV are presented in Table 18 for preference, and Table 19 for treatment. The full tables of results are presented in Appendix A.4.1, Table A 4 for preference and Table A 5 for treatment. For analysis of data by preference, comparisons were made to the Preferred treatment, and to the No-stimulus treatment for analysis by treatment.

Analysis by preference

Heart rate and IBI were affected similarly by preference. Compared to the Preferred treatment, average HR increased during the Aversive treatment by 10.3 bpm (25.9%), and IBI decreased by 284.3 ms (18.0%). HR and IBI did not differ between the Intermediate and Preferred treatments. The mean HRs for the Preferred, Intermediate, and Aversive treatments were 39.7 bpm, 39.9 bpm, and 49.9 bpm respectively.

Compared to the Preferred treatment, SDNN increased during the Aversive treatment by 8.0 ms (15.3%) but was unchanged during the Intermediate treatment. RMSSD tended to decrease by 3.9 ms (6.9%) during the Intermediate treatment ($p = 0.084$) but did not differ during the Aversive treatment. The ratio of RMSSD:SDNN decreased during the Aversive treatment by 0.15 (13.2%) but was not significantly affected by the Intermediate treatment.

Table 18 Statistically significant results from generalised least square regression models describing the relationship between five parameters for heart rate variability and preference for a treatment.
The intercept equates to the value during the Preferred treatment as there was only one variable in the model. Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean
LogHRav (bpm)	Intercept (Preferred)	3.68	0.05	<0.001	39.65
	Aversive	***0.23***[a]	***0.02***	***<0.001****	***49.90***
IBIav (ms)	Intercept (Preferred)	1575.88	76.58	<0.001	1575.88
	Aversive	***-284.30***	***21.28***	***<0.001****	***1291.58***
LogSDDN (ms)	Intercept (Preferred)	3.97	0.10	<0.001	51.98
	Aversive	**0.14**	**0.06**	**0.019***	**59.95**
LogRMSSD (ms)	Intercept (Preferred)	4.05	0.11	<0.001	56.40
	Intermediate	-0.07	0.04	0.084	52.52
Ratio RMSSD/SDNN	Intercept (Preferred)	1.14	0.03	<0.001	1.14
	Aversive	***-0.15***	***0.02***	***<0.001****	***0.99***

NS No statistically significant results
* p ≤ 0.05
[a] The logHRav for the aversive was 0.23 higher than during the preferred treatment. That is the predicted logHRav is 3.91 (3.68 + 0.23), which was statistically significant (p < 0.001).

Analysis by treatment

Heart rate and IBI were similarly affected by treatment. Compared to the No-stimulus treatment, the Groom treatment resulted in an increase in HR by 2.0 bpm (5.1%) and a decrease in IBI by 52.8 ms (3.3%). The Spray treatment resulted in an increase in HR of 11.0 bpm (28.4%) and decrease in IBI of 290.0 ms (18.3%). Heart rate and IBI did not differ between the Person and No-stimulus treatments. The mean HRs for the No-stimulus, Groom, Person, and Spray treatments were 38.9 bpm, 40.9 bpm, 38.9 bpm and 49.9 bpm respectively.

Compared to the No-stimulus treatment, SDNN increased during the Spray treatment by 11.2 ms (22.8%) but was unchanged during the Person and Groom stimuli. RMSSD did not vary by treatment. The ratio of RMSSD:SDNN decreased during the Spray treatment by 0.14 (12.4%) but was not significantly affected by the Person or Groom stimuli.

Table 19 Statistically significant results from generalised least square regression models describing the relationship between five parameters for heart rate variability and treatment.
The intercept equates to the value during the No-stimulus treatment as there was only one variable in the model. Analysis is based on comparison to the No-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean
LogHrav (bpm)	Intercept (No-stimulus)	3.66	0.05	<0.001	38.86
	Groom	***0.05*** [a]	***0.01***	***<0.001****	***40.85***
	Spray	***0.25***	***0.02***	***<0.001****	***49.90***
IBIav (ms)	Intercept (No-stimulus)	1581.08	76.58	<0.0001	1581.08
	Groom	***-52.84***	***17.60***	***0.003****	***1528.24***
	Spray	***-289.96***	***21.67***	***<0.001****	***1291.12***
LogSDDN (ms)	Intercept (No-stimulus)	3.91	0.1	<0.001	48.90
	Spray	***0.20***	***0.06***	***<0.001****	***60.07***
LogRMSSD (ms)	NS				
Ratio RMSSD/SDNN	Intercept (No-stimulus)	1.13	0.03	<0.001	1.13
	Spray	***-0.14***	***0.02***	***<0.001****	***0.99***

NS No statistically significant results
* $p \leq 0.05$
[a] The logHRav was 0.05 significantly higher while being Groomed than during the No-stimulus treatment ($p < 0.001$). The model predicts that logHRav for groom treatment is 3.71 (3.66 + 0.05).

7.3.2 Eye temperature

The observed data for average and maximum ET are presented in box plots in Appendix A.4.2. The statistically significant results of the analysis of average and maximum ET are presented in Table 20 for preference, and Table 21 for treatment. The full tables of results are presented in Appendix A.4.2, Table A 6 for preference and Table A 7 for treatment. For analysis of data by preference, comparisons were made to the Preferred treatment, and to the No-stimulus treatment for analysis by treatment. Air temperature was found to have a significant linear effect on eye temperature in all models, whereas the effect of humidity was not significant.

Data was missing for one replicate for one horse due to equipment malfunction.

Analysis by preference

After adjusting for ambient air temperature and clustering at the level of the horse and replicate, the average ET was 0.07°C lower during the Aversive treatment than when a horses' Preferred treatment was applied. There was no significant difference in average ET between the Intermediate and Preferred treatments. The results of the mixed effects models that accounted for ambient air temperature and the clustering at the level of the horse and replicate found no significant impact of preference on the maximum ET.

Table 20 Statistically significant results of a mixed effects regression model examining the effect of preference for a treatment on average and maximum eye temperature (°C).
Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean^
Average eye temperature	Intercept (Preferred)	36.05 [b]	0.17	<0.001	36.05
	Aversive	***-0.07*** [c]	***0.02***	***0.001****	***35.98***
	Initial ambient air temperature [a]	***0.08*** [d]	***0.02***	***0.004****	-
Maximum eye temperature	Intercept (Preferred)	36.29	0.17	<0.001	36.29
	Initial ambient air temperature [a]	***0.07***	***0.02***	***0.007****	-

^ Model predicted mean at 15°C air temperature
* p ≤ 0.05
[a] Ambient air temperature centred at 15°C.
[b] When the preferred treatment was applied and the ambient air temperature was 15°C, the predicted average eye temperature was 36.05°C.
[c] The average eye temperature during the aversive treatment was 0.07°C less than during the preferred treatment, that is the predicted average eye temperature at 15°C was 35.08°C (36.05-0.07). The difference was statistically significant (p = 0.001).
[d] The average eye temperature increased by 0.08°C with every degree increase in initial ambient air temperature. This effect may not be biologically relevant as ambient air temperature was only recorded at the start of each session.

Analysis by treatment

After adjusting for ambient air temperature and clustering at the level of the horse and replicate, and when compared to the No-stimulus treatment, the presence of a Person resulted in an increase in average ET of 0.04°C, and the Spray treatment resulted in a decrease of 0.06°C. There was no significant difference in average ET between the Groom and

No-stimulus treatments. The results of the mixed effects models that accounted for ambient air temperature and the clustering at the level of the horse and replicate found no significant impact of treatment on the maximum ET.

Table 21 Statistically significant results of a mixed effects regression model examining the effect of treatment on average eye temperature (°C). Analysis is based on comparison to the no-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean^
Average eye temperature	Intercept (No-stimulus)	36.04 [b]	0.17	<0.001	36.04
	Person	**0.04**	**0.02**	**0.025***	**36.08**
	Spray	***-0.06*** [c]	***0.02***	***0.003****	***35.98***
	Initial ambient air temperature [a]	***0.08*** [d]	***0.02***	***0.004****	-
Maximum eye temperature	Intercept (No-stimulus)	36.28	0.17	<0.001	36.28
	Initial ambient air temperature [a]	***0.07***	***0.02***	***0.007****	-

^ Model predicted mean at 15°C air temperature
* $p \leq 0.05$
[a] Ambient air temperature centred at 15°C.
[b] When the No-stimulus treatment was applied and ambient air temperature was 15°C, the average eye temperature was 36.04°C.
[c] The average eye temperature during the Spray treatment was 0.06°C less than during the No-stimulus treatment, that is the predicted average eye temperature during the Spray treatment at 15°C was 35.08°C (36.04-0.06). The difference was statistically significant ($p = 0.003$).
[d] The average eye temperature increased by 0.08°C with every degree increase in initial ambient air temperature. This effect may not be biologically relevant as ambient air temperature was only recorded at the start of each session.

7.4 Discussion

7.4.1 Heart rate variability and preference

The increase in heart rate, decrease in inter-beat interval, and decrease in RMSSD:SDNN ratio with the Aversive preference (or Spray treatment) were consistent with previous research (von Borell et al. 2007). They indicate a predominance of the sympathetic nervous system over the parasympathetic nervous system, which may be due to increased arousal and/or negative emotion.

The observed increase in SDNN with the Aversive/Spray treatment contradicts what might be expected based on the literature (Boissy et al. 2007b; von Borell et al. 2007; Zebunke et al. 2011). In the current study, it is possible that the intermittent nature of the Aversive/Spray stimulus and the recording length could have had an impact. Analysis was performed across the whole treatment period, which may have masked the decrease in SDNN that was observed within the first 80-100 seconds, which is when the Sprays were delivered. While outside the boundaries of this particular study, future research could investigate the duration of cardiac responses to emotion inducing stimuli and their return to normal values.

Alternatively, there may have been a spurious impact on SDNN and RMSSD from the results of one horse. Interestingly, no such aberrance was obvious in the inter-beat interval data for this horse. It is possible that a cardiac abnormality or other pathological state was present in this horse that was abnormally affecting heart rate variability, but not heart rate. Additionally, one horse was observed to have very little cardiac reaction to the Aversive/Spray treatment. It is difficult to select an aversive stimulus that is sufficiently aversive yet is safe to perform in restrained horses, is not painful, and is ethically acceptable.

The mean heart rate and heart rate variability parameters did not differ significantly between the Preferred and Intermediate treatments. It could be that the emotional valence did not differ across preferences/treatments, with only an increase in arousal with the Aversive/Spray treatment that was reflected by the increase in heart rate. However, this seems illogical given that in the previous experiment, the Preferred treatments were voluntarily chosen by the horses, and the Spray treatment was not preferred. Thus, if they all had the same emotional value, then no preference would have been seen. It may be that horses do not experience large reactions to non-aversive stimuli, thus differentiating small differences between emotionally positive and more neutral stimuli may require larger numbers of horses than used in this study.

There was a tendency toward a small increase in RMSSD with the Preferred treatments compared to the Intermediate ones, which may suggest higher parasympathetic nervous system activity and pleasure with the Preferred treatment (Visser et al. 2002; von Borell et al. 2007). Similarly, putatively pleasant experience of music or massage were associated with PNS predominance in horses (Kowalik et al. 2017; Wiśniewska et al. 2019). In contrast, no differences were found in HR and HRV in two studies involving grooming, wither scratch or neck patting with authors suggesting this may be due to high arousal, predominance of the SNS, and/or high inter-individual variability (Thorbergson et al. 2016; Lansade et al. 2018). It may be that RMSSD is confounded by arousal and is indicative of valence only in the lower arousal situations (without predominance of SNS), such as with the Intermediate and Preferred treatments. As the difference between means was small, the variability within the data may have overshadowed the true effects and repetition of the experiment with more subjects may elucidate a more significant effect.

7.4.2 Heart rate variability and treatment

The finding that heart rate increased with the Groom treatment is contrary to early research showing a reduction in heart rate with grooming/massage (Feh et al. 1993; McBride et al. 2004). However, support may be found in more recent studies that suggest an increase in arousal from stroking or brushing (Schmied et al. 2008; Janczarek et al. 2018b; Lansade et al. 2018; Tamioso et al. 2018). Horses displayed a significant reaction to stroking, with HRV indicating a shift towards SNS dominance, and the authors concluding that horses were more excited during stroking than when at rest, depending on the body region being stroked, type, and sex of the horse (Janczarek et al. 2018b). In another study, no significant effect on HR or HRV (SDNN, RMSSD) was found in horses that were groomed either pleasantly or aversively, although there was a tendency towards a lower SDNN (SNS predominance) with the aversive treatment (Lansade et al. 2018). The authors suggest that contrary to other studies on grooming/tactile stimulation in horses where horses were relaxed, horses in their study were aroused in both the positive and negative treatment.

It may be that in the current study, horses were more aroused during the Groom treatment than the No-stimulus one, and the methodological and analytical differences in the studies account for the contrasting findings. Alternatively, the response from horses that did not like the Groom treatment may have been larger in magnitude, and thus overshadowed the response from those that did, hence the importance of analysis by preference rather than by treatment.

Restraint of the horses in stocks limited the amount and type of physical activity possible, but the potential impact of physical movement on heart rate variability could not be completely eliminated (Lenoir et al. 2017). The effect of artefacts on data also needs to be considered, as a substantial effect on heart rate variability can be erroneously seen, particularly with automated error correction methods (von Borell et al. 2007). In this study, artefacts were replaced with missing values, rather than interpolated using the Polar software and the equipment manufacturer's undisclosed proprietary methods, which may not be appropriate for horses (Lanata et al. 2015). The corrected data was suitable for analysis (von Borell et al. 2007). Further research is required to examine the effect of different methods of artefact determination and data processing on heart rate variability outcomes (Lanata et al. 2015; Thorbergson et al. 2016). Future research should report details on the method of error correction.

Humans were present during this experiment, and two of the treatments, Groom and Person, directly involved humans. An effect on HRV has been seen during emotional transfer via chemosignals from humans to horses (Lanata et al. 2018). Horses' autonomic responses, as measured by HRV, differed between exposure to the control, human fearful, and human happy sweat samples, with a higher response in the fear odour condition. There are implications of these findings for the conduct of studies on animals, and the possible confounding effects of human emotion.

It may be interesting, in future research, to compare results from other indices of heart rate and heart rate variability, such as area under the curve, frequency domain, and geometric means, to those of commonly used time domain measures, with reference to positive emotional stimuli.

7.4.3 Eye temperature

The average temperature of the eye and surrounding tissue was found to vary significantly with the stimuli used in this study. However, in contrast to previous research, maximum eye temperature was not found to differ between treatment categories (Stubsjoen et al. 2009; Stewart et al. 2010b). The decrease in average eye temperature with the Aversive/Spray treatment is supported by Stewart et al. (2008b) and likely reflects vasoconstriction mediated by stimulation of the sympathetic nervous system over the acute phase (Stewart et al. 2008a). At the time the experiment was conducted, this was the first finding of a decrease in eye temperature with an unpleasant stimulus in horses.

The lower magnitude of eye temperature response in the current study may be due to use of a comparatively less aversive stimulus than electric prod or disbudding (Stewart et al. 2008b; Stewart et al. 2008a). An effect of lesser magnitude might also be expected when looking at an average rather than a maximum value. Additionally, eye temperature was averaged over the entire Spray treatment period (180 seconds) to maintain consistency across treatments. Which may have also lowered the apparent differences between means as the Spray was only delivered during the first 40 seconds of the period. Further research is required to examine the duration of effect of emotion on physiological parameters.

In contrast to the current study, an increase in eye temperature in horses' has been weakly associated with stress (Bartolomé et al. 2013), fear (Dai et al. 2015), and controversially, with a putatively aversive procedure (McGreevy et al. 2012). It is possible that the direction of eye temperature change is species and/or stimulus specific. However, eye temperature has been seen to increase above baseline following an acute drop due to pain, fear or stress in cattle (Stewart et al. 2008a). The contrasting findings may be due to methodological differences, particularly around the timing of stimulus application and eye temperature measurement.

The increase in eye temperature with the Person treatment may be due to vasodilation mediated by stimulation of the parasympathetic nervous system. This could indicate a pleasant emotion and may warrant further investigation. It is also possible that the response is stimulus specific, as horses were observed attempting to engage with the person, perhaps increasing local/regional blood flow and temperature due to movement.

Initial ambient air temperature was included in the analysis models and an effect was found. As air temperature increased above 15^{o}C, horses' average eye temperature increased by 0.08^{o}C. The effect of a 1^{o}C change in air temperature was greater than the effect of treatment or preference. It should be noted that the IRT camera automatically calibrates for ambient temperature (and humidity, which was not found to have a significant effect during analysis), intermittently during use. However, the effect found during analysis may have been incorrectly estimated, as the ambient temperature was recorded at the start of each session, not continuously. Nevertheless, it suggests that the observed preference/treatment differences were within physiological range.

7.4.4 Conclusion

In conclusion, heart rate, heart rate variability, and average eye temperature varied with preference for a treatment and varied with some treatments. Based on the heart rate variability and eye temperature results, the sympathetic nervous system predominated in the aversive condition, as expected. It may be that heart rate and measures of sympathetic nervous system predominance are indicative of emotional arousal (Iyasere et al. 2022). In contrast with earlier studies in horses, which showed a reductive impact of stroking on heart rate, suggesting a calming effect, there was no conclusive evidence of a relaxing effect or parasympathetic nervous system predominance with grooming in the current study.

Further research is required to explore the tendency for RMSSD to increase with positive emotion induced by a preferred stimulus, when compared to an intermediately preferred one, and if this is evident only when the sympathetic nervous system does not predominate. Future research could focus on the duration of physiological responses following emotion inducing stimuli.

Chapter 8 Body behaviours as indicators of emotion in horses

8.1 Introduction

In humans, body postures are indicative of emotional states (Keeling 2019). In animals, body posture can be used to infer discomfort or pain, and may inform about emotion or intent. Specific body parts, such as head, neck, and tail position or movement, may be useful as indicators of emotion in animals (Keeling 2019; Kremer et al. 2020). Behavioural measures are also likely to be important indicators of positive emotion in animals (Whittaker et al. 2019). However, none are definitive, most may be affected by arousal level, and they may be context dependent.

The same behaviour may be observed in different contexts, with different underlying emotions or motivations. Behaviours may be species-specific and impacted by many factors, including breed/strain, age, sex, reproductive status, social status, and altered mental health (e.g., abnormal repetitive or stereotypic behaviour). Communicative functions of behaviour may require the presence of a suitable audience in order to be displayed. Behaviours also need to occur often enough, and/or for long enough, to enable useful analysis.

Behaviour measurement can be influenced subjectively (Whittaker et al. 2019). There are a variety of methods of measuring behaviour in quantitative and semi-quantitative ways which can impact the data and conclusions of a study (Martin et al. 2007). For example, the movement of individual body parts (e.g., hind leg lift, head/neck raise), whole body position (e.g., standing, lying), or activity (e.g., walking, eating, drinking), can be recorded from direct observations, coded from audio/video recordings, or inferred from positioning data such as tri-axial accelerometry. There are variations in sampling frame, for example all occurrences, focal animal sampling, and time sampling. Behaviours may be recorded and analysed as binary (observed at least once/not observed), frequency (number of times or rate of occurrence), or as a duration (or proportion) of time data. Behavioural motivation or intent can be indirectly assessed, for example, by measuring approach/avoidance behaviour e.g., distance, latency to approach, and distance over time. Behaviour can also be scored using a graded scale, which may involve interpretation of motivation or intent e.g., threat behaviour patterns which may consist of individual behavioural units such as ears back, teeth bared, rump presentation, and/or kick threat. There are advantages and disadvantages of each method and data type. It may be easier and quicker to score patterns of behaviour, perform time sampling, or infer motivation. However, there may be more interpretation and subjectivity involved, along with decreases in repeatability and/or validity, and restrictions on analysis due to ordinal or categorical data type. Measuring all occurrences of individual postures or movements may be more objective. Interpretation of behaviour is complex and requires a cautious approach to avoid bias (Whittaker et al. 2019).

Behaviours require careful cataloguing in an ethogram, and pilot studies may be needed to refine descriptions and recording or coding methods. Existing ethograms for behaviour in horses have been used as a foundation for the development of a study specific ethogram for use in this experiment (McDonnell et al. 1995; McDonnell et al. 2002; McDonnell 2003).

The aim of this experiment was to determine whether vocalisations, elimination behaviour, or neck, limb, or tail movements varied with preference for a treatment (Preferred, Intermediate, or Aversive), or varied with the treatment (Groom, Person, Spray, or No-stimulus).

8.2 Materials and methods

Video recordings taken from the side view camera (Figure 23; Sony DCR-SR85E, Auckland, New Zealand) were analysed using Interact software (version 9, Mangold International GmbH, Arnstorf, Germany) to record all occurrences and duration of behaviours pertaining to the body of the horse. The same observer viewed all the videos. The ethogram for body behaviours and vocalisations is presented in Table 22 (modified from McDonnell 2003).

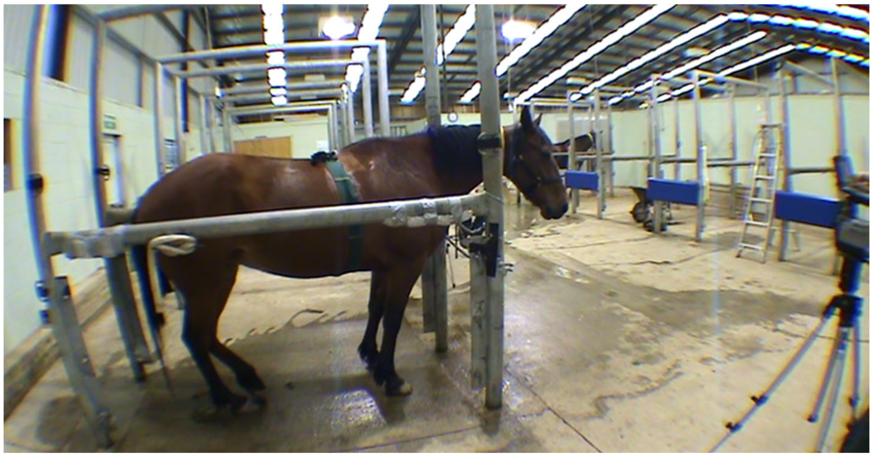


Figure 23 View of the horse as recorded by the side camera with a fisheye lens. The behaviours observable in this particular image are Hindleg rest, Neck neutral, and Tail neutral.

Table 22 Ethogram of body behaviours recorded from video. Exclusions are behaviours which are incompatible with each behaviour, for example, a horse could not be said to lift its hindleg and kick at the same time.

Body part	Behaviour	Exclusions	Description
Foreleg	Lift	All other foreleg	Complete lift of foreleg hoof from ground in a vertical direction
	Strike	All other foreleg	Complete lift of foreleg hoof from ground also with leg movement forward
	Rear*	All other foreleg	Both front legs lift from ground
Hindleg	Lift	All other hindleg	Complete lift of hindleg hoof from ground in a vertical direction
	Kick	All other hindleg	Complete lift of hindleg hoof from ground surface also with leg movement backwards
	Rest	All other hindleg	Heel bulb of hindleg removed from ground while toe of hoof remains in contact with ground
	Stand*	All other hindleg	Soles of both hooves in full contact with ground
Neck position	Very low	All other neck positions	Neck lowered approximately -30 to -90º from horizontal through vertebral column
	Low	All other neck positions	Neck lowered approximately -10 to -30º from horizontal through vertebral column
	Neutral	All other neck positions	Neck is horizontal with the vertebral column +/-10º
	High	All other neck positions	Neck approximately 10-45º above the horizontal through vertebral column

Body part	Behaviour	Exclusions	Description
	Very high	All other neck positions	Neck approximately 45-90º above horizontal through vertebral column
Tail	Lift	All other tail	Movement of dock in vertical plane resulting in dock placement higher than horizontal to spine/pelvis - including tail lift before elimination
	Swish	All other tail	Lateral movement of the dock without moving in a vertical plane away from the hindquarters
	Thrash	All other tail	Strong repeated vertical dock movement from the hindquarters with a gap clearly visible between the dock and hindquarters, with or without lateral movement.
	Neutral*	All other tail	Tail held in typical position, neither clamped nor raised
Vocalisation	Snort	Other vocalisations	Loud, harsh sound made by quick forceful movement of air
	Whinny (neigh)	Other vocalisations	Loud, prolonged call, beginning high pitched, ending lower pitched
	Nicker*	Other vocalisations	Low pitched, gutturally pulsated sound
Elimination	Urination*	All leg positions	Horse urinates, leg position not counted during urination
	Defecation	All leg positions	Horse defecates, leg position not counted during defecation
Other	Out of view*	Everything	Horse not able to be seen on video. To allow for accurate proportion of time calculations.

* Data were not subsequently analysed as either included for calculation of proportion of time or rate, were neutral positions, or did not occur often enough to analyse.

8.2.1 Data Analysis

The statistical methods used depended on whether the outcome was analysed as a binary, count, or linear variable.

The following behaviours were not often observed, and as such, analysis was based on whether or not the event occurred (binary): fore strike, hind kick, tail lift, tail thrash, snort, whinny, any vocalisation, and defecation. The relationship between preference or treatment and occurrence of these behaviours was modelled separately using a mixed effects logistic regression model (GLMM – logistic).

The following behaviours occurred more frequently, and so analysis focused on the number of times an event happened during the period (count): fore lift, hind lift, hind rest, neck very low, neck very high, and tail swish. The relationship between preference or treatment and number of times these behaviours were observed was modelled separately using a mixed effects Poisson (log-linear) model (GLMM – Poisson), with an off-set for the duration of each period.

The following behaviours occurred for a reasonable duration of time, and so analysis focused on the proportion of time for each period that the horse was observed in this state (linear): neck low, neck neutral, and neck high. The relationship between the proportion of time spent in these states and preference or treatment was modelled separately using a mixed linear regression model (LMM).

The following behaviours did not occur often enough across all treatments to be analysed: urination, rear, and nicker. There was insufficient variation in tail neutral and changes from the neutral position for tail were analysed as above, thus tail neutral was not analysed separately.

An odds ratio (OR) was given for the binary data, a rate ratio (RR) for the count data, and beta for the proportion of time data. Model predicted values were also given. For binary data, this represented the probability of the behaviour occurring, for count data it was the predicted count for a three-minute treatment period, and for behaviour duration data it was the predicted proportion of time the behaviour occurred for.

Intra-observer reliability for the behaviour data was assessed by comparison of results from a repeated view of randomly chosen horses/replicates, and calculation of the Pearson's correlation between the repeated observations.

When examining the proportion of time spent performing behaviours, where only two behavioural options are possible, for example a hoof touching the ground or not, it may not be necessary to report results for both behaviours as one can be deduced from the other (once time out of view has been accounted for). However, where more than two behaviours are possible, for example neck position very low, low, neutral, high, very high, it cannot be assumed which alternate behaviour is being performed, and thus necessary to report the full results. Similarly, for binary data, it may not be necessary to report the number of observations wherein the behaviour did not occur. In contrast, when measuring counts of behaviours, the frequency of one behaviour does not provide information about performance of another, so it is necessary to report the results for all behaviours.

8.3 Results

There were 21 behaviours described in the ethogram (excluding the out of view code), which were coded, and for which the binary, count, and duration data was explored. Of these, five behaviours did not occur often enough across all treatments to enable them to be analysed, or were neutral behaviours: rear, hindleg stand, tail neutral, nicker, and urination. A new binary variable, ‘Any vocalisation’, was created, which was an amalgamation of snort, whinny, and nicker. In total, 17 variables, representing 16 behaviours, were analysed.

The average Pearson’s correlation between the repeated observations for the body behaviours was 0.99 (SE 0.004), and ranged from 0.86 to 1, which demonstrated very good intra-observer reliability (Petrie et al. 2013).

The statistically significant results of the analysis of body behaviours are presented in Table 23 for preference, and Table 24 for treatment. The full tables of results are presented in Appendix A.5.7, Table A 19 for preference and Table A 20 for treatment. For analysis of data by preference, comparisons were made to the Preferred treatment, and to the No-stimulus treatment for analysis by treatment.

8.3.1 Analysis by Preference

The Intermediate preference resulted in a decrease in the rate of neck very high (0.6 times), compared to the Preferred treatment. There were tendencies towards an increase in neck very low rate (1.2 times, $p = 0.066$) and a decrease in tail swish rate (0.8 times, $p = 0.095$).

The Aversive treatment resulted in an increase in the following behaviours: foreleg lift rate (2.2 times), hindleg lift rate (3.6 times), hindleg kick probability (6.8 times), hindleg rest rate (2.0 times), neck very low rate (1.4 times), neck high duration (1.8 times), neck very high rate (1.8 times), tail swish rate (3.5 times), and tail thrash probability (4.0 times), compared to the Preferred treatment. There was a decrease in the neck low duration (0.6 times) and neck neutral duration (0.8 times) behaviours.

8.3.2 Analysis by Treatment

The Groom treatment resulted in an increase in neck very low rate (1.2 times) and very high rate (1.6 times), compared to the No-stimulus treatment. There were tendencies towards statistical significance for increases in hind lift rate (1.2 times, $p = 0.063$), hindleg kick probability (8.7 times, $p = 0.086$), neck low duration (1.3 times, $p = 0.094$), and tail swish rate (1.4 times, $p = 0.078$), and a decrease in foreleg strike probability (0.2 times, $p = 0.094$).

The Person treatment resulted in an increase in neck low duration (1.5 times) and tail swish rate (1.5 times) behaviours, and decreases in hindleg lift rate (0.8 times), neck high duration (0.8 times), tail lift probability (0.4 times), any vocalisation probability (0.3 times), and defecation probability (0.1 times) behaviours. There was a tendency towards a statistically significant decrease in snort probability (0.3 times, $p = 0.066$).

The Spray treatment resulted in an increase in the following behaviours: foreleg lift rate (2.2 times), hindleg lift rate (3.3 times), hindleg kick probability (15.1 times), hindleg rest rate (2.2 times), neck very low rate (1.4 times), neck high duration (1.6 times), neck very high rate (3.1 times), tail swish rate (5.3 times), and tail thrash probability (6.5 times), compared to the No-stimulus treatment. There was a decrease in neck neutral duration (0.8 times).

Table 23 Statistically significant results from models examining the effect of preference for a treatment on the occurrence, frequency, or proportion of time spent displaying various body behaviours.
Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Foreleg	Lift count [a]	Preferred	REF				3.1
		Aversive	***0.79 (0.08)***	***2.21 (1.89-2.57)*** [d]	**-**	***<0.001****	***6.9***
	Strike [b^]	NS					
Hindleg	Lift count [a]	Preferred	REF				2.1
		Aversive	***1.28 (0.08)***	***3.60 (3.06-4.23)***	**-**	***<0.001****	***7.7***
	Kick [b^]	Preferred	REF				0.6%
		Aversive	**1.92 (0.90)**	**-**	**6.81 (1.18-39.49)** [e]	**0.032***	**4.0%**
	Rest count [a]	Preferred	REF				2.0
		Aversive	***0.69 (0.11)***	***1.99 (1.60-2.48)***	**-**	***<0.001****	***3.9***
Neck	Very low count [a]	Preferred	REF				2.4
		Intermediate	0.14 (0.08)	1.15 (0.99-1.34)	-	0.066	2.8
		Aversive	***0.34 (0.10)***	***1.41 (1.16-1.72)***	**-**	***<0.001****	***3.4***
	Low time [c]	Intercept (Preferred)	0.11 (0.03)	-	-	<0.001	11.1%
		Aversive	**-0.05 (0.02)** [f]	**-**	**-**	**0.015***	**6.3%**
	Neutral time [c]	Intercept (Preferred)	0.60 (0.05)	-	-	<0.001	59.7
		Aversive	***-0.14 (0.04)***	**-**	**-**	***<0.001****	***46.0***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	High time [c]	Intercept (Preferred)	0.25 (0.05)	-	-	<0.001	24.6%
		Aversive	***0.19 (0.04)***	***-***	***-***	***<0.001****	***43.6%***
	Very high count [a]	Preferred	REF				0.2
		Intermediate	***-0.52 (0.18)***	***0.59 (0.42-0.84)***	***-***	***0.004****	***0.1***
		Aversive	***0.58 (0.19)***	***1.78 (1.23-2.59)***	***-***	***0.002****	***0.3***
Tail	Lift [b^]	NS					
	Swish count [a]	Preferred	REF				0.8
		Intermediate	-0.25 (0.15)	0.78 (0.58-1.04)	-	0.095	0.6
		Aversive	***1.24 (0.15)***	***3.47 (2.60-4.63)***	***-***	***<0.001****	***2.9***
	Thrash [b^]	Preferred	REF				4.1%
		Aversive	**1.38 (0.69)**	**-**	**3.98 (1.03-15.33)**	**0.045***	**14.7%**
Vocalisation	Snort [b^]	NS					
	Whinny [b^]	NS					
	Any [b^]	NS					
Elimination	Defecation [b^]	NS					

NS No statistically significant results
* $p \leq 0.05$
^ Caution with interpretation as large SE.
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the Preferred treatment, the rate of foreleg lift behaviour was 2.21 (95%CI 1.89 - 2.57) times greater during the Aversive treatment and this was highly statistically significant ($p < 0.001$). Foreleg lift was predicted to occur 6.9 times during the three-minute Aversive treatment.
[e] Compared to the Preferred treatment, the odds of a hindleg kick occurring was 6.81 (95%CI 1.18 - 39.49) times higher during the Aversive treatment and this was statistically significant ($p = 0.032$). There was a 4.0% predicted probability of occurrence of a hindleg kick during the Aversive treatment. Due to a large SE of the beta, the confidence interval around the odds ratio is wide, and caution with interpretation is advised.
[f] Compared to the Preferred treatment, the proportion of time that horses had their neck in the low/very low position was 0.05 (SE 0.02) less during the Aversive treatment, this was statistically significant ($p = 0.015$). The predicted percentage of time that horses had their neck in the low position was 11% for the Preferred treatment and 6% for the Aversive treatment

Table 24 Statistically significant results from models examining the effect of treatment on the occurrence, frequency or proportion of time spent displaying various body behaviours.
Analysis is based on comparison to the no-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Foreleg	Lift count [a]	No-stimulus	REF				3.1
		Spray	***0.79 (0.08)***	***2.21 (1.90-2.58)***	***-***	***<0.001****	***6.9***
	Strike [b^]	No-stimulus	REF				0.1%
		Groom	-1.74 (1.04)	-	0.18 (0.02-1.34)	0.094	0.03%
Hindleg	Lift count [a]	No-stimulus	REF				2.3
		Person	**-0.20 (0.10)**	**0.82 (0.68-0.99)** [d]	**-**	**0.036***	**1.9**
		Spray	***1.20 (0.08)***	***3.33 (2.84-3.90)***	***-***	***<0.001****	***7.7***
	Kick [b^]	No-stimulus	REF				0.3%
		Groom	2.16 (1.26)	-	8.67 (0.73-102.39)	0.086	2.1%
		Spray	**2.72 (1.21)**	**-**	**15.12 (1.41-162.61)**	**0.025***	**3.6%**
	Rest count [a]	No-stimulus	REF				1.8
		Spray	***0.78 (0.11)***	***2.18 (1.74-2.73)***	***-***	***<0.001****	***3.9***
Neck	Very low count [a]	No-stimulus	REF				2.5
		Groom	**0.18 (0.09)**	**1.20 (1.01-1.42)**	**-**	**0.037***	**2.9**
		Spray	***0.32 (0.10)***	***1.38 (1.13-1.68)***	***-***	***0.001****	***3.4***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Low time [c]	Intercept (No-stimulus)	0.09 (0.03)	-	-	0.004	9.0%
		Groom	0.03 (0.02)	-	-	0.094	11.6%
		Person	***0.05 (0.02)***	***-***	***-***	***0.003****	***13.8%***
	Neutral time [c]	Intercept (No-stimulus)	0.60 (0.05)	-	-	<0.001	59.9
		Spray	***-0.14 (0.04)***	***-***	***-***	***<0.001****	***46.0***
	High time [c]	Intercept (No-stimulus)	0.27 (0.05)	-	-	<0.001	26.5%
		Person	**-0.06 (0.03)**	**-**	**-**	**0.031***	**20.1%**
		Spray	***0.17 (0.04)*** [e]	***-***	***-***	***<0.001****	***43.5%***
	Very high count [a]	No-stimulus	REF				0.1
		Groom	**0.46 (0.22)**	**1.58 (1.02-2.45)**	**-**	**0.039***	**0.2**
		Spray	***1.14 (0.22)***	***3.14 (2.03-4.86)***	***-***	***<0.001****	***0.3***
Tail	Lift [b^]	No-stimulus	REF				34.3%
		Person	**-0.87 (0.42)**	**-**	**0.42 (0.18-0.96)**	**0.040***	**18.0%**
	Swish count [a]	No-stimulus	REF				0.5
		Groom	0.34 (0.19)	1.40 (0.96-2.03)	-	0.078	0.8
		Person	**0.41 (0.19)**	**1.51 (1.04-2.17)**	**-**	**0.029***	**0.8**
		Spray	***1.67 (0.17)***	***5.30 (3.79-7.40)***	***-***	***<0.001****	***2.9***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Thrash [b^]	No-stimulus	REF				2.5%
		Spray	**1.88 (0.75)**	**-**	**6.54 (1.51-28.39)**	**0.012***	**14.4%**
Vocalisation	Snort [b^]	No-stimulus	REF				14.8%
		Person	-1.09 (0.59)	-	0.34 (0.11-1.07)	0.066	5.5%
	Whinny [b^]	NS					
	Any [b^]	No-stimulus	REF				21.4%
		Person	**-1.08 (0.51)**	**-**	**0.34 (0.13-0.92) [f]**	**0.033***	**8.5%**
Elimination	Defecation [b^]	No-stimulus	REF				8.9%
		Person	**-2.30 (1.10)**	**-**	**0.10 (0.01-0.86)**	**0.036***	**1.0%**

NS No statistically significant results
* $p \leq 0.05$
^ Caution with interpretation as large SE.
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the No-stimulus treatment, the rate of hindleg lift behaviour was 0.82 (95%CI 0.68 - 0.99) times less during the Person treatment, this was statistically significant (p = 0.036). Hindleg lift was predicted to occur 1.9 times during the three-minute Person treatment.
[e] Compared to the No-stimulus treatment, the proportion of time that horses had their neck in the high position was 0.17 (SE 0.04) more during the Spray treatment, this was statistically significant (p < 0.001). The predicted percentage of time that horses had their neck high/very high position was 27% for the No-stimulus treatment and 44% for the Spray treatment.
[f] Compared to the No-stimulus treatment, the odds of any vocalisation occurring was 0.34 (95%CI 0.13 - 0.92) times lower during the Person treatment and this was statistically significant (p = 0.033). There was an 8.5% predicted probability of occurrence of a vocalisation during the Person treatment.

8.4 Discussion

Few body behaviours showed differences between the Preferred and Intermediate conditions. More body behaviours differed, and there was a larger magnitude of effect, between the Preferred (positive) and Aversive (negative) emotional experiences. Most studies compare a single positive to a single negative emotional condition. The inclusion of three levels of emotional experience in this study allows suggestions about causality to be made. This finding supports the notion that most of the impact on behaviour is due to negative emotion, rather positive emotion, and confirms that positive emotion produces more subtle effects.

8.4.1 Body behaviour and preference

Neck behaviour was analysed in the current study as the frequency of different positions and the time spent above, at, or below the neutral position. High neck positions, or head height, in horses is associated with increased arousal and/or attention to the environment or a specific stimulus, and low positions with low arousal and/or relaxation (McGreevy 2004). Horses experiencing a Preferred stimulus moved their necks to the very high position more than with an Intermediate, but less than with an Aversive stimulus. It may be that horses were more attentive to a more meaningful stimulus (Preferred versus Intermediate), without the hypervigilance/anxiety provoked by an Aversive stimulus. This is supported by other studies suggesting that putatively pleasant stimuli, such as stroking or grooming, are emotionally arousing in horses and sheep (Janczarek et al. 2018b; Lansade et al. 2018; Tamioso et al. 2018). The frequency of the neck in a very high position may be an indicator of emotional experience in horses.

Contrary to expectation, horses experiencing a Preferred stimulus tended to move their neck to the very low position fewer times than with the Intermediate or Aversive treatments. Thus, the neck very low position may not necessarily be a sign of positive emotion, at least in the context of this study.

There is support from the literature for the decrease in duration of neck neutral, increase in rate of neck very high and increase in duration of neck above neutral behaviours observed with the Aversive experience (Lansade et al. 2018). Low head position at two time points was associated with increased fear and stress in cattle and sheep at an abattoir (Hemsworth et al.

2011). This aligns with the results of the current experiment, where horses experiencing the Aversive stimulus moved their neck to the very low position more times but held it below neutral for a decreased proportion of the time, compared to the Preferred one. An increase in neck-head movements to extreme positions and a preponderance for longer time in neck high position may be indicative of high arousal, negative emotion.

In sheep tail wagging is suggested to be an indicator of positive emotion in response to voluntary brushing by a human (Tamioso et al. 2018). In contrast, in the current study, tail swish behaviour was markedly increased with the Aversive relative to the Preferred treatment, but tended to be moderately increased with the Preferred relative to the Intermediate treatment. The rate of tail swish behaviour in horses is an indicator of emotional experience but the relationship of tail movement and emotion may be species specific.

Most of the other changes seen with the Aversive treatment, for example, the increase in fore and hind leg lifts, and hindleg kick, were not counter-intuitive. They may represent negative emotional valence, increased emotional arousal, and/or a motivation to escape from, or fight, a perceived threat (McGreevy 2004).

In contrast, the frequency of hindleg rest, a behaviour which is usually associated with resting or drowsing in horses (McGreevy 2012), was increased in horses experiencing the Aversive treatment, along with hindleg kick occurrence and hindleg lift count. It may be that in the context of an unpleasant stimulus, the hind leg is rested/pointed in readiness for further movement (lift, kick). This is supported in the ethogram of Alert behaviour by (McDonnell et al. 1995), where the drawing depicts a horse with a hind leg at rest.

8.4.2 Body behaviour and treatment

When compared to the No-stimulus treatment, there were two behaviours which varied across the other three treatments. They were hindleg lift and tail swish rates. The No-stimulus treatment had the lowest rate of tail swishing, which might indicate low arousal. Interestingly, hindleg lifts decreased with the Person treatment and tended to increase with the Groom treatment. This may reflect a difference in agitation or arousal. Alternatively, it may be due to the position of the treatment delivery, with the person in front distracting attention forwards, whereas both the Spray and Groom treatments were delivered at the side of the horse. The Person treatment also allowed some control by the horse. The horse was able to choose to investigate the person, or ignore them, whereas there was no controllability for the horse during the Groom or Spray treatments. It may be that hindleg lifts reflect agitation due to frustration from lack of control.

The nature of a particular treatment may impact behaviour independent of emotion. For example, the time spent with neck below neutral was increased, and the time above neutral tended to be decreased, with the Person treatment. These results could be interpreted as signs of low arousal or relaxed pleasure, but also could have been due to horses reaching out to investigate the person, and may also have been influenced by the height of the person. A moderate increase in the neck very high behaviour was found with the Groom treatment, relative to the No-stimulus treatment, which was not apparent with the Person treatment. Horses were observed to extend their head and neck upwards during the Groom treatment, perhaps in attempt at mutual grooming (McGreevy 2004), and was similarly noted in a study by Lansade et al. (2018). Although some consideration should be given to context dependency and stimulus specificity, these findings highlight the value of interpretation of results by preference rather than by treatment.

It is also possible that the behaviour changes seen with the Spray treatment may be specific to the type of stimulus. That is, in this study, it cannot be ruled out that negative emotion (fear or anxiety) caused by a sudden, unpredictable, novel, intermittent stimulus, in this case the Spray, might cause a different behavioural reaction than another type of aversive stimulus, for example, pain from an injection.

A few behaviours were observed rarely, or only in a small number of horses. The decision was made not to analyse some of the behaviours, such as in the case of nickering, which was observed four times in total with four horses displaying it. In other cases, the behaviour was considered further in statistical models (e.g., whinny, tail thrash). When interpreting the results from events that occurred rarely, it is important to note that non-significance should not be viewed as evidence that the behaviour is not linked to emotion, as the power to detect significance would be low. Given that these behaviours tended to occur in only a small number of horses, future studies investigating less frequently observed behaviours should include a larger number of horses, rather than increased replicates in a smaller number of horses.

8.4.3 Conclusion

The frequency of neck movements to extreme positions, tail lift, swish, and thrash, foreleg lift, hindleg lift, kick, and rest, defecation, and duration with neck above, at, or below neutral, varied by preference and/or treatment. The treatment analysis findings emphasise the value of analysis of results by preference, or another method of independent identification of an individual's emotional regard for the treatments.

Negative emotional experience has more impact on body behaviour than positive emotional experience. Inclusion of three levels of emotional experience is useful. Analysis of the duration and frequency of a behaviour, for example neck height, may yield more insight than either data type alone.

In the context of this study, negative emotional experience is indicated by an increased rate of neck very low or neck very high position, tail swishing, fore or hind leg lifts, longer time with neck above neutral, or an increased likelihood of a tail thrash or hindleg kick occurring. Positive emotional experience is associated with a moderately increased rate of neck very high position, and possibly a decreased rate of neck very low position or a mildly increased rate of tail swishing.

It is acknowledged that the findings may be species specific, and/or reflect display of behavioural intention, and that some behaviours may be impacted by specific treatments. Further research is required to corroborate and expand on these results. Future research may involve further use of remote monitored devices and/or artificial intelligence to remove any potential impact of human influence and improve efficiency of data collection and interpretation.

Chapter 9 Head and facial movements as indicators of emotion in horses

9.1 Introduction

Facial movement behaviours, which together form characteristic facial expressions, may function to convey information to conspecifics, and possibly other species, about emotion, motivation, and/or intent (Lezama-García et al. 2019; Whittaker et al. 2019). Facial movement behaviours are mainly voluntarily expressed, and may be suppressed in some situations, for example prey species hiding signs of weakness (Lezama-García et al. 2019; van Dierendonck et al. 2020). However, in humans, facial expressions are less able to be suppressed than other indicators of emotion and may provide a more accurate assessment of emotion than gross behaviours (Descovich et al. 2017).

Most emotion research involving head and facial movements has been conducted in animals experiencing negative emotion, such as pain (Flecknell 2010; Langford et al. 2010; Leach et al. 2011; Sotocinal et al. 2011; Dalla Costa et al. 2014; Keeling 2019; Whittaker et al. 2019; Mogil et al. 2020). It has enabled facial expression to be regarded as a valid scientific measure in animals. More recently, facial movement has been explored for indicators of positive emotion in animals (Keeling 2019; Whittaker et al. 2019; Mogil et al. 2020).

The display and appearance of facial movement behaviours may be impacted by many factors, including anatomy, facial morphology, the nature of a particular stimulus (for example prehension of food), vocalisation, and the presence and nature of an audience, which may include humans (Paul et al. 2005; Guesgen et al. 2017).

It is reasonable to suggest that facial behaviours may have communicative functions in horses (Lansade et al. 2018). Horses are a highly social species requiring well-developed communication mechanisms. They rely on visual information, and have complex facial muscles, allowing for a wide range of facial movements (Wathan et al. 2015). The superficial facial musculature of horses includes a concentration of muscles around the base of the ears, the eyes, lips, nares, and cheeks, enabling a wide range of possible facial movements.

Previous work on emotional experience in other species suggests that ear position, and/or rate of movement, ear skin colour, eye aperture, visible sclera (eye white), nares, mouth, jaw, tongue, and lips, may be important in display of emotion (Sandem et al. 2002; Proctor et al. 2015; Descovich et al. 2017). In horses, ear position, eyebrow wrinkles or lift, eyelid twitches, rate of eye blinks, and non-masticatory chewing, may provide information on emotional experience (Hintze et al. 2016), although there have been contrasting results (Whittaker et al. 2019).

The aim of this experiment was to determine whether head behaviours and facial movements varied with preference for a treatment (Preferred, Intermediate, or Aversive), or varied with the treatment (Groom, Person, Spray, or No-stimulus).

9.2 Materials and methods

High definition video recordings taken with the head camera (Figure 24; Canon Legria HFM31, Sydney, Australia) were analysed using Interact software (version 9, Mangold International GmbH, Arnstorf, Germany) to record all occurrences and duration of behaviours pertaining to the head of the horse, including ear and facial movements. The same observer viewed all the videos. The ethogram for head and facial behaviours is presented in Table 25. The ethogram was based on previous literature (McDonnell et al. 1995; McDonnell et al. 2002; McDonnell 2003; Reefmann et al. 2009b; Reefmann et al. 2009a).

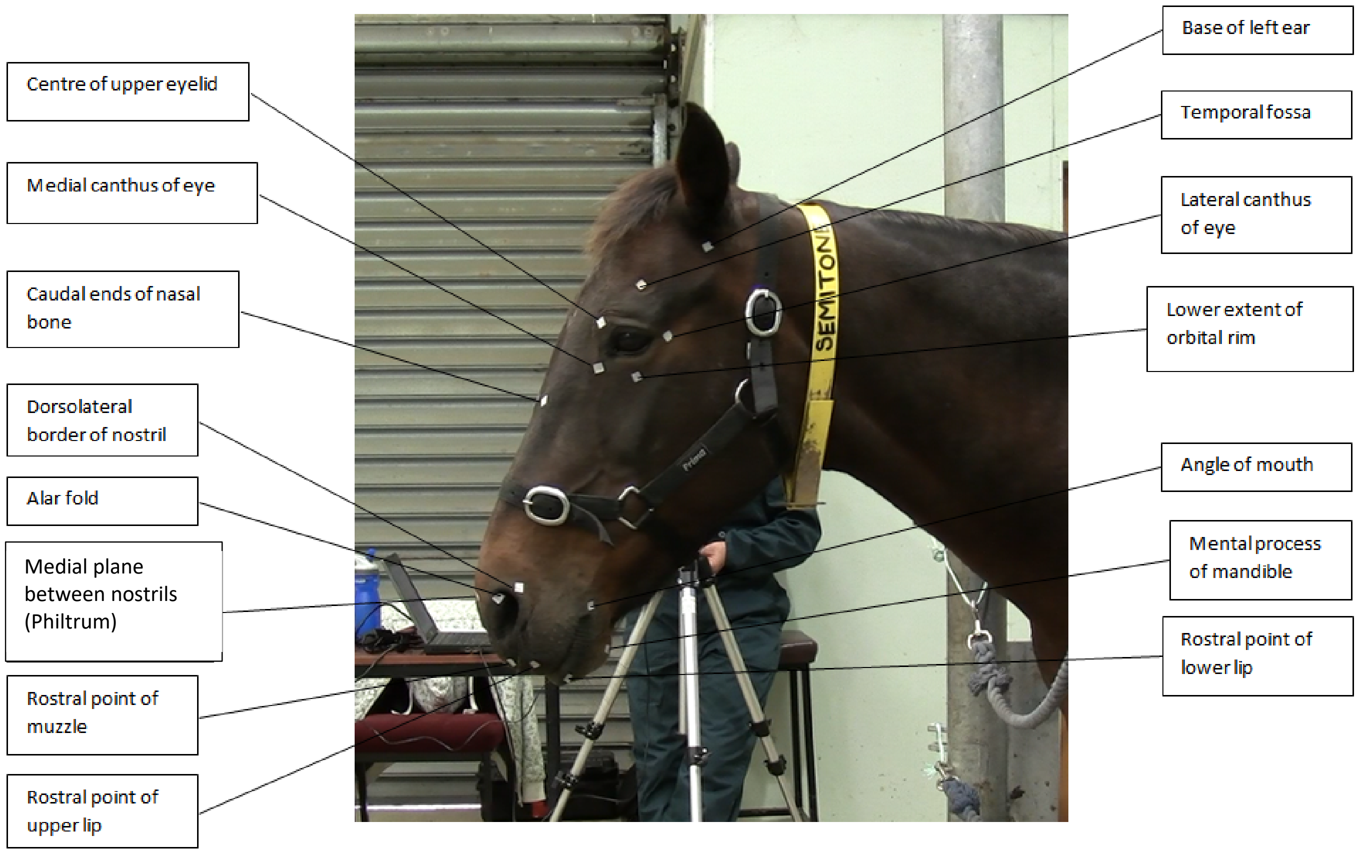


Figure 24 View of the horse as recorded by the head camera.
The positions of white markers (1 cm x 1 cm flexible, breathable self-adhesive dressing; Fixomull, Smith and Nephew, Sydney, Australia) on prominent head features are shown.

Table 25 Ethogram of head and facial behaviours recorded from video. Exclusions are behaviours which were incompatible with each behaviour, for example, a horse could not be said to be lip licking and chewing at the same time it was yawning, eating, or performing oral investigation.

Head part	Behaviour	Exclusions	Description
Head position	Left^	Right, centre, weaving	Poll turned to horse's left of body in horizontal plane
	Centre^	Right, left, weaving	Poll central i.e., within lateral extremities of body
	Right^	Left, centre, weaving	Poll turned to horse's right of body in horizontal plane
	Rotation^	Shake	Head tilted around cervical vertebrae in vertical plane
	Shake	Rotation	Repeated head rotation around cervical vertebrae
	Extended^	Neutral, tucked, tossing	Chin extended away from neck. Chin-throat angle > 90°
	Neutral^	Extended, tucked, tossing	Chin-throat angle neutral, approx. 90°
	Tucked^	Extended, neutral, tossing	Chin tucked into neck. Chin-throat angle decreased < 90°
	Tossing	Extended, neutral, tucked	Repeated, rapid, forceful vertical flicking of head/nose
	Weaving*	Left, right, centre	Repeated lateral movements of head from one side to the other
Ears	Left forward	All other left ear	Left ear rotated forward
	Right forward	All other right ear	Right ear rotated forward
	Left side	All other left ear	Left ear rotated to left side
	Right side	All other right ear	Right ear rotated to left side
	Left back	All other left ear	Left ear rotated backward but meatus still visible
	Right back	All other right ear	Right ear rotated backward but meatus still visible
	Left flat	All other left ear	Left ear flat back against neck, meatus flattened or not visible
	Right flat	All other right ear	Right ear flat back against neck, meatus flattened or not visible
	Left flicking*	All other left ear	Rapid movement of left ear back and forth
	Right flicking*	All other right ear	Rapid movement of right ear back and forth

Head part	Behaviour	Exclusions	Description
Eye	Blink	Sclera	Momentary eyelid closure
	Sclera	Blink	Sclera/white of eye visible
	Aperture small	All other aperture	Small area of eyeball visible. Narrowed palpebral fissure
	Aperture neutral	All other aperture	Eye aperture intermediate, neither large nor small
	Aperture large	All other aperture	Large area of eyeball visible. Widened palpebral fissure
	Eyebrow wrinkle		Wrinkling of skin above eyebrow
	Eyebrow angle forward#	All other eyebrow angle	Position of eyebrow angle medial to centre of eye
	Eyebrow angle mid#	All other eyebrow angle	Position of eyebrow angle at centre of eye
	Eyebrow angle back#	All other eyebrow angle	Position of eyebrow angle lateral to centre of eye
	Temporal fossa convex^	Temporal fossa flat	Fossa bulging
	Temporal fossa flat^	Temporal fossa convex	Fossa flat
Mouth	Contracted lips	Eating, yawn, vocal, lip lick/chew, investigation	Decreased distance from commissures of lips to rostral point of lips, pursed lips.
	Drooping lip	Apposing lips, top lip extended	Lower lip not apposing upper lip and projects forward
	Top lip extended^	Apposing lips, drooping lip	Top lip extends forward of bottom lip
	Lip licking & chewing	Eating, investigation, yawn	Tongue flicked over lips or chewing movements with mouth. Excluding when access to food
	Face pulling	All other mouth	Extended or repetitive contortion of face that does not fit into other discrete categories
	Oral investigation	Eating, yawn, lip licking/chewing	Sniffing, mouthing (grasping with lips or teeth) an object or person. Excluding when access to food.
	Yawn*	Eating, investigation, vocalisation	Large opening of mouth
	Eating*	Lip licking/chewing, yawn, investigation	Chewing, lip licking with food in mouth, prehending food

Head part	Behaviour	Exclusions	Description
Chin	Chin wobble	All other chin	Repeated, rapid movement of chin excluding other mouth movements (yawn, eating, lip licking/chewing, investigation)
	Parrot mouth*	All other chin, all mouth	Upper and lower lips puckered. Upper lip extends beyond lower lip similar to a parrot's beak. Excluding other mouth movements (yawn, eating, lip licking/chewing, investigation)
	Puckered chin	All other chin	Mentum more pointed, skin puckers under chin
Cheek	Cheek muscle contraction*	Eating, lip licking/chewing	Masseter muscle moving. Excluding when access to food, when licking and chewing, vocalising, or yawning
	Cheek concave^	All other cheek	Cheek blown out
	Cheek taut/convex^	All other cheek	Cheek sucked in or taut
Nares	Flared	All other nares	Increased distance between medial and lateral alar fold
	Neutral	All other nares	Nares in typical position neither narrowed nor flared
	Narrowed	All other nares	Nares narrowed or flattened
Out of view	Out of view right ear*	All right ear	Cannot determine right ear position as not in camera view
	Out of view left ear*	All left ear	Cannot determine left ear position as not in camera view
	Out of view eye*	All eye	Cannot determine eye behaviour as not in camera view
	Out of view mouth*	All mouth	Mouth/chin behaviour cannot be determined as out of camera view
	Out of view nares*	All nares	Nares behaviour cannot be determined as out of camera view

^ Coding was ceased for behaviours that were too infrequent, unable to be detected reliably, or were co-dependent.
* Data were not subsequently analysed as either included for calculation of proportion of time or rate or did not occur often enough to analyse.
[#] Due to difficulty distinguishing between the position of the eyebrow angle, the eyebrow angle forward, mid, and back behaviours were amalgamated during coding into a single behaviour termed 'Eyebrow angle'.

9.2.1 Data Analysis

The statistical methods used depended on whether the outcome was to be analysed as a binary, count, or linear variable.

The following behaviours were not often observed and, as such, analysis was based on whether or not the event occurred (binary): head shake, head toss, left ear flat, right ear flat, large eye aperture, contracted lips, drooping lip, face pulling, puckered chin, and nares narrowed. The relationship between preference or treatment, and occurrence of these behaviours, was modelled separately using a mixed effects logistic regression model (GLMM - logistic).

The following behaviours occurred more frequently, and analysis focused on the number of times an event happened during the period (count): left ear forward, right ear forward, left ear side, right ear side, left ear back, right ear back, eye blink, eye sclera visible, small eye aperture, neutral eye aperture, eyebrow wrinkle, eyebrow angle, lip licking or chewing, oral investigation, chin wobble, nares flared, and nares neutral. The relationship between preference or treatment, and number of times these behaviours were observed, was modelled separately using a mixed effects Poisson (log-linear) model (GLMM – Poisson), with an off-set for the duration of each period.

The following behaviours occurred for a reasonable duration of time, and analysis focused on the proportion of time that the horse was observed in this state for each period (linear): left ear forward, right ear forward, left ear side, right ear side, left ear back, right ear back, and nares flared. When calculating the proportion of time that a behaviour was displayed for, the period length was adjusted for the time that the relevant head part was out of view. The relationship between the proportion of time spent in these states, and preference or treatment, was modelled separately using a mixed linear regression model (LMM).

An odds ratio (OR) was given for the binary data, a rate ratio (RR) for the count data, and beta for the proportion of time data. Model predicted values were also given; for binary data, this represented the probability of the behaviour occurring, for count data it was the predicted count for a three-minute treatment period, and for behaviour duration data it was the predicted proportion of time the behaviour occurred for.

Intra-observer reliability for the behaviour data was assessed by comparison of results from a repeated view of randomly chosen horses/replicates, and calculation of the Pearson's correlation between the repeated observations.

9.3 Results

There were 48 behaviours described in the ethogram (excluding the five out of view codes). Following observation of recordings from two horses, the number of behaviours coded was reduced. Coding was ceased for behaviours that were infrequent, if observations were not reliable (e.g., due to subtlety of changes or relative camera angle), or if the performance was dependent on another behaviour (e.g., head tucked/extended with investigation). On this basis, 12 behaviours were not coded: head left, centre, and right, head rotation, head extended, neutral, and tucked, temporal fossa convex and flat, top lip extended, and cheek concave and convex. The three eyebrow angle behaviours, forward, mid, and back, were amalgamated during coding into one variable: eyebrow angle.

The binary, count, and time data for 34 different behaviours were explored. Seven behaviours did not occur often enough across all treatments to enable them to be analysed: head weaving, left and right ear flicking, yawn, eating, parrot mouth, and cheek contraction. The count and duration data were analysed for seven behaviours: left and right ears forward, left and right ears side, left and right ears back, and nares flared. In total, 34 variables, representing 27 behaviours, were analysed.

Due to missing files, the head video was substituted with footage from the oblique camera for parts of two horse replicates. The results are not expected to be impacted as the video quality was sufficient, and it was used for a small proportion of the data collection (4/231 periods, i.e., 1.7%).

The average Pearson's correlation between the repeated observations for the head behaviours was 0.96 (SE 0.009), and ranged from 0.83 to 1, which demonstrated very good intra-observer reliability (Petrie et al. 2013).

The statistically significant results of the analysis of head and facial behaviours are presented in Table 26 for preference, and Table 27 for treatment. The full tables of results are presented in Appendix A.6.1, Table A 38 for preference and Table A 39 for treatment. For analysis of data by preference, comparisons were made to the Preferred treatment, and to the No-stimulus treatment for analysis by treatment.

9.3.1 *Analysis by Preference*

Intermediate versus Preferred treatments

The Intermediate preference resulted in an increase in the left ear forward rate (1.1 times), right ear forward rate (1.1 times), left ear back rate (1.1 times), right ear back rate (1.2 times), blink rate (1.1 times), eyebrow angled rate (1.3 times), and nares flared rate (1.1 times), compared to the Preferred treatment. There were tendencies towards statistical significance for an increase in the right ear side rate (1.1 times, $p = 0.058$), and nares neutral rate (1.1 times, $p = 0.062$).

The Intermediate preference resulted in a decrease in small eye aperture rate (0.6 times), contracted lips probability (0.3 times), oral investigation probability (0.7 times), and chin wobble rate (0.9 times), compared to the Preferred treatment.

Aversive versus Preferred treatments

The Aversive treatment resulted in an increase in the following behaviours: left ear forward rate (1.2 times), left ear back rate (1.8 times), right ear back rate (1.7 times), left ear back duration (1.7 times), right ear back duration (1.8 times), blink rate (1.2 times), sclera visible rate (1.3 times), neutral eye aperture rate (1.4 times), large eye aperture probability (28.0 times), eyebrow wrinkle rate (1.3 times), eyebrow angled rate (1.9 times), contracted lips probability (11.7 times), nares flared rate (2.3 times), and nares flared duration (1.6 times), compared to the Preferred treatment.

The Aversive treatment resulted in a decrease in the following behaviours: right ear forward rate (0.9 times), left ear forward duration (0.8 times), right ear forward duration (0.5 times), right ear side rate (0.9 times), left ear side duration (0.5 times), right ear side duration (0.6 times), small eye aperture count (0.4 times), drooping lower lip probability (0.2 times), lip licking and chewing rate (0.5 times), oral investigation probability (0.2 times), and chin wobble rate (0.6 times), compared to the Preferred treatment. There was a tendency towards statistical significance for a decrease in the left ear side count (0.9 times, $p = 0.057$).

9.3.2 Analysis by Treatment

Groom versus No-stimulus treatments

The Groom treatment resulted in an increase in the following behaviours: head shake probability (2.9 times), blink rate (1.1 times), eyebrow angle rate (1.5 times), lip licking and chewing rate (1.2 times), and oral investigation rate (2.0 times), compared to the No-stimulus treatment. There were tendencies towards statistical significance for an increase in left ear flat probability (7.8 times, $p = 0.091$), right ear flat probability (7.4 times, $p = 0.095$), and chin pucker probability (8.3 times, $p = 0.071$).

The Groom treatment resulted in a decrease in the left ear back rate (0.9 times), left ear back duration (0.8 times), and the small eye aperture rate (0.6 times), compared to the No-stimulus treatment. There was a tendency towards statistical significance for a decrease in the left ear forward rate (0.9 times, $p = 0.074$).

Person versus No-stimulus treatments

The Person treatment resulted in an increase the following behaviours: left ear flat probability (10.3 times), right ear flat probability (9.7 times), small eye aperture rate (1.5 times), oral investigation count (2.0 times), and nares flared rate (1.2 times), compared to the No-stimulus treatment. There were tendencies towards statistical significance for an increase in the contracted lips probability (3.7 times, $p = 0.070$), and nares neutral rate (1.1 times, $p = 0.072$).

The Person treatment resulted in a decrease in the following behaviours: left ear forward rate (0.9 times), right ear forward rate (0.9 times), right ear forward duration (0.7 times), and right ear side count (0.9 times), compared to the No-stimulus treatment. There were tendencies towards statistical significance for a decrease in the left ear forward duration (0.9 times, $p = 0.071$), and blink rate (0.95 times, $p = 0.076$).

Spray versus No-stimulus treatments

The Spray treatment resulted in an increase in the following behaviours: left ear back rate (1.6 times), right ear back rate (1.5 times), left ear back duration (1.7 times), right ear back duration (1.8 times), left ear flat probability (16.4 times), right ear flat probability (21.8 times), blink rate (1.1 times), sclera visible rate (1.3 times), aperture neutral rate (1.6 times), large eye aperture probability (61.5 times), eyebrow angled rate (1.9 times), contracted lips probability (37.4 times), nares flared rate (2.2 times), nares flared duration (1.7 times), and nares neutral rate (1.9 times), compared to the No-stimulus treatment.

The Spray treatment resulted in a decrease in the following behaviours: right ear forward rate (0.9 times), left ear forward duration (0.7 times), right ear forward duration (0.5 times), left ear side rate (0.9 times), right ear side rate (0.8 times), left ear side duration (0.5 times), right ear side duration (0.6 times), drooping lower lip probability (0.2 times), lip licking and chewing rate (0.6 times), oral investigation rate (0.6 times), and chin wobble rate (0.6 times), compared to the No-stimulus treatment. There was a tendency towards statistical significance for a decrease in the small eye aperture rate (0.3 times, $p = 0.061$).

Table 26 Statistically significant results from models examining the effect of preference for a treatment on the occurrence, frequency, or proportion of time spent displaying various head behaviours.
Analysis is based on comparison to the Preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Head	Shake [b^]	NS					
	Tossing [b^]	Preferred	REF				8.2%
		Aversive	1.10 (0.61)	-	2.99 (0.91-9.80)	0.070	21.1%
Ears	Left forward count [a]	Preferred	REF				12.7
		Intermediate	***0.13 (0.04)***	***1.13 (1.05-1.22)***	**-**	***<0.001****	***14.4***
		Aversive	***0.20 (0.05)***	***1.23 (1.11-1.36)*** [e]	**-**	***<0.001****	***15.6***
	Right forward count [a]	Preferred	REF				14.2
		Intermediate	**0.09 (0.04)**	**1.10 (1.02-1.18)**	**-**	**0.012***	**15.6**
		Aversive	***-0.15 (0.06)***	***0.86 (0.77-0.97)***	**-**	***0.010****	***12.3***
	Left forward time [c]	Intercept (Preferred)	0.25 (0.05)	-	-	<0.001	25.4%
		Aversive	**-0.06 (0.03)** [f]	**-**	**-**	**0.017***	**19.8%**
	Right forward time [c]	Intercept (Preferred)	0.23 (0.04)	-	-	<0.001	23.0%
		Aversive	***-0.11 (0.03)***	**-**	**-**	***<0.001****	***11.8%***
	Left side count [a]	Preferred	REF				17.9
		Aversive	**-0.10 (0.05)**	**0.91 (0.82-1.00)**	**-**	**0.057***	**16.3**

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right side count [a]	Preferred	REF				17.7
		Intermediate	**0.07 (0.03)**	**1.07 (1.00-1.14)**	**-**	**0.058***	**18.9**
		Aversive	**-0.12 (0.05)**	**0.88 (0.80-0.98)**	**-**	**0.022***	**15.6**
	Left side time [c]	Intercept (Preferred)	0.37 (0.04)	-	-	<0.001	36.7%
		Aversive	***-0.18 (0.03)***	***-***	***-***	***<0.001****	***18.6%***
	Right side time [c]	Intercept (Preferred)	0.37 (0.04)	-	-	<0.001	36.6%
		Aversive	***-0.16 (0.03)***	***-***	***-***	***<0.001****	***20.4%***
	Left back count [a]	Preferred	REF				9.9
		Intermediate	**0.11 (0.04)**	**1.12 (1.02-1.22)**	**-**	**0.012***	**11.0**
		Aversive	***0.58 (0.05)***	***1.78 (1.60-1.98)***	***-***	***<0.001****	***17.6***
	Right back count [a]	Preferred	REF				10.5
		Intermediate	***0.19 (0.04)***	***1.21 (1.11-1.31)***	***-***	***<0.001****	***12.7***
		Aversive	***0.54 (0.05)***	***1.72 (1.55-1.91)***	***-***	***<0.001****	***18.1***
	Left back time [c]	Intercept (Preferred)	0.35 (0.07)	-	-	<0.001	35.4%
		Aversive	***0.26 (0.04)***	***-***	***-***	***<0.001****	***61.1%***
	Right back time [c]	Intercept - Preferred	0.38 (0.06)	-	-	<0.001	38.2%
		Aversive	***0.29 (0.04)***	***-***	***-***	***<0.001****	***67.2%***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Left flat [b^]	NS					
	Right flat [b^]	NS					
Eye	Blink count [a]	Preferred	REF				41.1
		Intermediate	***0.06 (0.02)***	***1.07 (1.02-1.12)***	-	***0.008****	***43.8***
		Aversive	***0.16 (0.03)***	***1.17 (1.10-1.25)***	-	***<0.001****	***48.1***
	Sclera count [a]	Preferred	REF				3.9
		Aversive	***0.26 (0.08)***	***1.30 (1.10-1.53)***	-	***0.002****	***5.1***
	Aperture small count [a]	Preferred	REF				0.9
		Intermediate	***-0.58 (0.14)***	***0.56 (0.42-0.74)***	-	***<0.001***	***0.5***
		Aversive	***-0.85 (0.25)***	***0.43 (0.26-0.70)***	-	***<0.001***	***0.4***
	Aperture neutral count [a]	Preferred	REF				6.8
		Aversive	***0.35 (0.07)***	***1.42 (1.23-1.64)***	-	***<0.001****	***9.6***
	Aperture large [b^]	Preferred	REF				1.7%
		Aversive	***3.33 (0.90)***	-	***27.97 (4.83-162.15)***	***<0.001****	***32.6%***
	Eyebrow wrinkle count [a]	Preferred	REF				1.7
		Aversive	**0.27 (0.14)**	**1.31 (1.00-1.72)**	-	**0.050***	**2.2**

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Eyebrow angle count [a]	Preferred	REF				1.4
		Intermediate	**0.24 (0.12)**	**1.27 (1.01-1.61)**	**-**	**0.039***	**1.8**
		Aversive	***0.64 (0.14)***	***1.89 (1.43-2.50)***	**-**	***<0.001****	***2.7***
Mouth	Contracted lips [b^]	Preferred	REF				8.3%
		Intermediate	**-1.14 (0.59)**	**-**	**0.32 (0.10-1.00)**	**0.051***	**2.8%**
		Aversive	***2.46 (0.61)***	**-**	***11.74 (3.57-38.55)***	***<0.001****	***51.5%***
	Drooping lip [b^]	Preferred	REF				33.1%
		Aversive	***-1.74 (0.68)***	**-**	***0.18 (0.05-0.66)***	***0.010****	***8.0%***
	Lip licking & chewing count [a]	Preferred	REF				4.0
		Aversive	***-0.65 (0.13)***	***0.52 (0.41-0.67)***	**-**	***<0.001****	***2.1***
	Face pulling [b^]	NS					
	Oral investigation count [a]	Preferred	REF				1.2
		Intermediate	***-0.35 (0.11)***	***0.71 (0.57-0.87)***	**-**	***0.001****	***0.8***
		Aversive	***-1.45 (0.25)***	***0.23 (0.14-0.38)***	**-**	***<0.001****	***0.3***
Chin	Chin wobble count [a]	Preferred	REF				4.0
		Intermediate	**-0.16 (0.08)**	**0.85 (0.73-0.99)**	**-**	**0.035***	**3.4**
		Aversive	***-0.54 (0.13)***	***0.58 (0.45-0.74)***	**-**	***<0.001****	***2.3***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Puckered chin [b^]	NS					
Nares	Flared count [a]	Preferred	REF				6.3
		Intermediate	**0.11 (0.06)**	**1.12 (1.00-1.25)**	**-**	**0.053***	**7.0**
		Aversive	***0.82 (0.07)***	***2.27 (2.00-2.59)***	***-***	***<0.001****	***14.2***
	Flared time [c]	Intercept (Preferred)	0.04 (0.01)	-	-	<0.001	4.3%
		Aversive	***0.03 (0.01)***	***-***	***-***	***<0.001****	***7.0%***
	Neutral count [a]	Preferred	REF				14.0
		Intermediate	0.08 (0.04)	1.08 (1.00-1.17)	-	0.062	15.1
		Aversive	***0.68 (0.05)***	***1.98 (1.79-2.18)***	***-***	***<0.001****	***27.6***
	Narrowed [b^]	NS					

NS No statistically significant results
* $p \leq 0.05$
^ Caution with interpretation as large SE
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the Preferred treatment, the odds of a head shake occurring was 0.45 (95%CI 0.19-1.03) times lower during the Aversive treatment and this was not statistically significant ($p = 0.059$). There was an 11.0% predicted probability of occurrence of a head shake during the Aversive treatment.
[e] Compared to the Preferred treatment, the rate of left ear forward behaviour was 1.23 (95%CI 1.11 - 1.36) times greater during the Aversive treatment and this was statistically significant ($p < 0.001$). Left ear forward was predicted to occur 15.6 times during the three-minute Aversive treatment.
[f] Compared to the Preferred treatment, the proportion of time that horses had their left ear forward was 0.06 (SE 0.03) less during the Aversive treatment, this was statistically significant ($p = 0.017$). The predicted percentage of time that horses had their left ear forward was 25.4% for the Preferred treatment and 19.4% for the Aversive treatment.

Table 27 Statistically significant results from models examining the effect of treatment on the occurrence, frequency, or proportion of time spent displaying various head behaviours.
Analysis is based on comparison to the No-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Head	Shake [b^]	No-stimulus	REF				18.4%
		Groom	***1.08 (0.42)***	***-***	***2.94 (1.29-6.68)***	***0.010****	***39.4%***
	Tossing [b^]	NS					
Ears	Left forward count [a]	No-stimulus	REF				14.7
		Groom	-0.08 (0.04)	0.93 (0.85-1.01)	-	0.074	13.7
		Person	***-0.12 (0.04)***	***0.89 (0.81-0.96)*** [e]	***-***	***0.005****	***13.1***
	Right forward count [a]	No-stimulus	REF				15.7
		Person	***-0.15 (0.04)***	***0.86 (0.79-0.94)***	***-***	***<0.001****	***13.5***
		Spray	***-0.25 (0.06)***	***0.78 (0.70-0.87)***	***-***	***<0.001****	***12.3***
	Left forward time [c]	Intercept (No-stimulus)	0.28 (0.05)	-	-	<0.001	27.9%
		Person	-0.04 (0.02)	-	-	0.071	24.2%
		Spray	***-0.09 (0.02)*** [f]	***-***	***-***	***<0.001****	***19.4%***
	Right forward time [c]	Intercept (No-stimulus)	0.26 (0.04)	-	-	<0.001	26.2%
		Person	***-0.07 (0.02)***	***-***	***-***	***0.001****	***19.6%***
		Spray	***-0.14 (0.02)***	***-***	***-***	***<0.001****	***11.8%***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Left side count [a]	No-stimulus	REF				18.0
		Spray	**-0.10 (0.05)**	**0.90 (0.82-1.00)**	**-**	**0.044***	**16.3**
	Right side count [a]	No-stimulus	REF				19.1
		Person	***-0.08 (0.04)***	***0.92 (0.85-1.00)***	**-**	***0.040****	***17.6***
		Spray	***-0.20 (0.05)***	***0.82 (0.74-0.91)***	**-**	***<0.001****	***15.6***
	Left side time [c]	Intercept (No-stimulus)	0.36 (0.04)	-	-	<0.001	35.8%
		Spray	***-0.17 (0.03)***	**-**	**-**	***<0.001****	***18.6%***
	Right side time [c]	Intercept (No-stimulus)	0.37 (0.04)	-	-	<0.001	36.6%
		Spray	***-0.16 (0.03)***	**-**	**-**	***<0.001****	***20.4%***
	Left back count [a]	No-stimulus	REF				10.7
		Groom	**-0.10 (0.05)**	**0.91 (0.82-1.00)**	**-**	**0.053***	**9.7**
		Spray	***0.50 (0.05)***	***1.64 (1.48-1.82)***	**-**	***<0.001****	***17.6***
	Right back count [a]	No-stimulus	REF				12.3
		Groom	-0.09 (0.05)	0.92 (0.83-1.01)	-	0.068	11.3
		Spray	***0.38 (0.05)***	***1.47 (1.32-1.63)***	**-**	***<0.001****	***18.1***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Left back time [c]	Intercept (No-stimulus)	0.36 (0.07)	-	-	<0.001	35.7%
		Groom	**-0.06 (0.03)**	**-**	**-**	**0.050***	**29.8%**
		Spray	***0.25 (0.04)***	***-***	***-***	***<0.001****	***61.1%***
	Right back time [c]	Intercept (No-stimulus)	0.37 (0.06)	-	-	<0.001	37.1%
		Spray	***0.30 (0.04)***	***-***	***-***	***<0.001****	***67.2%***
	Left flat [b^]	No-stimulus	REF				0.3%
		Groom	2.06 (1.22)	-	7.81 (0.72-84.72)	0.091	2.4%
		Person	**2.34 (1.21)**	**-**	**10.33 (0.97-110.46)**	**0.053***	**3.1%**
		Spray	**2.80 (1.29)**	**-**	**16.43 (1.32-205.10)**	**0.030***	**4.9%**
	Right flat [b^]	No-stimulus	REF				0.4%
		Groom	2.01 (1.20)	-	7.44 (0.71-78.49)	0.095	2.9%
		Person	**2.27 (1.19)**	**-**	**9.72 (0.94-100.91)**	**0.057***	**3.8%**
		Spray	**2.80 (1.26)**	**-**	**21.78 (1.86-255.38)**	**0.014***	**8.1%**
Eye	Blink count [a]	No-stimulus	REF				42.1
		Groom	***0.10 (0.03)***	***1.11 (1.05-1.17)***	***-***	***<0.001****	***46.5***
		Person	-0.05 (0.03)	0.95 (0.90-1.01)	-	0.076	40.0
		Spray	***0.13 (0.03)***	***1.14 (1.07-1.22)***	***-***	***<0.001****	***48.1***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Sclera count [a]	No-stimulus	REF				3.8
		Spray	***0.28 (0.09)***	***1.32 (1.12-1.56)***	***-***	***<0.001****	***5.1***
	Aperture small count [a]	No-stimulus	REF				0.6
		Groom	***-0.54 (0.20)***	***0.58 (0.40-0.86)***	***-***	***0.007****	***0.4***
		Person	**0.39 (0.16)**	**1.48 (1.08-2.03)**	**-**	**0.014***	**0.9**
		Spray	-0.49 (0.26)	0.61 (0.37-1.02)	-	0.061	0.4
	Aperture neutral count [a]	No-stimulus					6.1
		Spray	***0.45 (0.07)***	***1.57 (1.35-1.81)***	***-***	***<0.001****	***9.6***
	Aperture large [b^]	No-stimulus	REF				0.8%
		Spray	***4.12 (1.04)***	***-***	***61.49 (7.98-474.01)***	***<0.001****	***32.4%***
	Eyebrow wrinkle count [a]	NS					
	Eyebrow angle count [a]	No-stimulus	REF				1.4
		Groom	***0.43 (0.13)***	***1.53 (1.19-1.97)***	***-***	***<0.001****	***2.1***
		Spray	***0.66 (0.14)***	***1.93 (1.46-2.56)***	***-***	***<0.001****	***2.7***
Mouth	Contracted lips [b^]	No-stimulus	REF				2.8%
		Person	1.30 (0.72)	-	3.67 (0.90-14.96)	0.070	9.4%
		Spray	***3.62 (0.78)***	***-***	***37.43 (8.17-171.53)***	***<0.001****	***51.5%***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Drooping lip [b^]	No-stimulus	REF				32.7%
		Spray	***-1.77 (0.69)***	***-***	***0.17 (0.04-0.65)***	***0.010****	***7.6%***
	Lip licking & chewing count [a]	No-stimulus	REF				3.8
		Groom	**0.19 (0.08)**	**1.20 (1.03-1.41)**	**-**	**0.023***	**4.6**
		Spray	***-0.61 (0.13)***	***0.55 (0.42-0.70)***	***-***	***<0.001****	***2.1***
	Face pulling [b^]	NS					
	Oral investigation count [a]	No-stimulus	REF				0.5
		Groom	***0.71 (0.15)***	***2.03 (1.50-2.74)***	***-***	***<0.001****	***1.0***
		Person	***0.95 (0.15)***	***2.60 (1.94-3.48)***	***-***	***<0.001****	***1.3***
		Spray	**-0.60 (0.27)**	**0.55 (0.32-0.93)**	**-**	**0.026***	**0.3**
Chin	Chin wobble count [a]	No-stimulus	REF				3.6
		Spray	***-0.45 (0.13)***	***0.64 (0.50-0.82)***	***-***	***<0.001****	***2.3***
	Puckered chin [b^]	No-stimulus	REF				0.0
		Groom	2.12 (1.17)	-	8.31 (0.83-83.02)	0.071	0.1
Nares	Flared count [a]	No-stimulus	REF				6.4
		Person	**0.14 (0.07)**	**1.15 (1.01-1.31)**	**-**	**0.036***	**7.4**
		Spray	***0.79 (0.07)***	***2.20 (1.94-2.51)***	***-***	***<0.001****	***14.2***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Flared time [c]	Intercept (No-stimulus)	0.04 (0.01)	-	-	<0.001	4.2%
		Spray	***0.03 (0.01)***	***-***	***-***	***<0.001****	***7.0%***
	Neutral count [a]	No-stimulus	REF				14.6
		Person	0.09 (0.05)	1.09 (0.99-1.20)	-	0.072	15.9
		Spray	***0.64 (0.05)***	***1.89 (1.72-2.09)***	***-***	***<0.001****	***27.6***
	Narrowed [b^]	NS					

NS No statistically significant results
* p ≤ 0.05
^ Caution with interpretation as SE may be large
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the No-stimulus treatment, the odds of a head shake occurring was 2.90 (95%CI 1.31 - 6.42) times higher during the Groom treatment and this was statistically significant (p = 0.009). There was a 39.5% predicted probability of occurrence of a head shake during the Groom treatment.
[e] Compared to the No-stimulus treatment, the rate of left ear forward behaviour was 0.89 (95%CI 0.81 - 0.96) times less during the Person treatment, this was statistically significant (p = 0.005). Left ear forward was predicted to occur 13.1 times during the three-minute Person treatment.
[f] Compared to the No-stimulus treatment, the proportion of time that horses had their left ear forward was 0.09 (SE 0.02) less during the Spray treatment, this was statistically significant (p < 0.001). The predicted percentage of time that horses had their left ear forward was 27.9% for the No-stimulus treatment and 19.4% for the Spray treatment.

9.4 Discussion

Previous work in other species suggested that ear position, eye aperture, nares and lips may indicate emotion (Reefmann et al. 2009b; Reefmann et al. 2009a; Langford et al. 2010; Sotocinal et al. 2011). As expected, the Aversive preference influenced more behaviours, and generally had a larger impact on the behaviours, than the Intermediate-Preferred comparisons, which were fewer and more subtle. The facial behaviour results are discussed below with reference to preference level, then by individual treatment. Due to the dearth of research considering the impact of a preferred/positive emotional experience compared to an intermediately preferred/less positive emotional experience (rather than an aversive/negative emotional experience), the findings for the non-aversive stimuli were initially considered together to enable comparison with the literature.

9.4.1 Facial behaviour and preference

The findings for a non-aversive stimulus (Preferred or Intermediate treatment) were consistent with the literature for small eye aperture in sheep (Reefmann et al. 2009a) and horses (Lansade et al. 2018). The rate of small eye aperture behaviour may be an indicator of emotion in horses and may be associated with more positive emotional experience.

Sclera visible behaviour was consistent with studies in cattle with purportedly low arousal positive valence (Sandem et al. 2002; Proctor et al. 2015; Sandem et al. 2006), and in horses (Hintze et al. 2016), but contrasted with other research in cattle involving eating a preferred food (Whittaker et al. 2019). It is possible that the increase in sclera visible in that study was due to a stimulus specific response and/or positioning of the stimulus. The sclera (eye white) is visible when the eye aperture is large and the head position and gaze are at extreme opposities (e.g. head left, gazing right, as in a sideways glance), or the eyeball is rotated in the orbit. It is thought to be associated with fear and submission or appeasement (Mills et al. 1999; McDonnell 2003; Sandem et al. 2004; Konig von Borstel et al. 2009). Thus, an increased sclera visible rate may indicate negative emotional experience in horses, and lower rate may indicate positive emotional experience.

During the Preferred treatment, there were more occurrences of chin wobbles or oral investigations than during the Aversive or the Intermediate treatments. There were also increased odds of a drooping lower lip compared to the Aversive condition. There is a paucity of comparable research, but it is interesting that these oral behaviours are grouped and positively relate to positive emotional experience. There were less odds of contracted lips than with the Aversive stimulus. Although research is required to better understand the function of the behaviour in horses, contraction or pursing of the lips is anecdotally associated with negative emotions such as confusion, disbelief, and frustration in humans, and may reflect generally increased muscular tension seen with negative experiences. Further research is required to corroborate the results which suggest that oral behaviours such as drooping lower lip, oral investigation, and chin wobbles are indicators of positive emotional experience.

Lip licking and chewing was seen more often in horses experiencing a non-aversive stimulus, and less often during the Spray treatment. Performance of non-nutritive lip licking and chewing (vacuum chewing) in horses has been said anecdotally to indicate submission, appeasement, relaxation, or evidence of learning, and is emphasised in various training approaches (Thorbergson et al. 2016; Hausberger et al. 2021). However, research suggests that it may be a displacement or affiliative behaviour, that also occurs during anticipation of feed (Schilder et al. 1984; Warren-Smith et al. 2007; Warren-Smith et al. 2008; McGreevy 2012). It is also suggested to occur after a stressful event is over, to aid rewetting of a mouth that is dry from sympathetic nervous system activation, and/or to relieve jaw tension after being tightly clenched (Warren-Smith et al. 2007; Warren-Smith et al. 2008; McGreevy 2012; Lie et al. 2018; Hausberger et al. 2021). It was found to be unrelated to learning or submission, and reduced with high arousal (heart rate > 100bpm) (Warren-Smith et al. 2007; Warren-Smith et al. 2008), which somewhat supports the results seen with the Aversive treatment in the current study.

Similar to the current study, Briefer et al. (2015b) suggested that chewing is associated with positive experiences. In contrast, its occurrence was suggested to be associated with a negative experience in horses, although there was no difference compared to the control situation (Thorbergson et al. 2016). In that study, behaviours were predetermined to reflect agitation or relaxation, or were termed ambiguous. There was no independent verification of how individual horses regarded the treatments. Importantly, horses were being ridden with bitted or bitless bridles which may impact oral behaviours differently compared to unbridled, unridden horses. Interpretation of lip licking and chewing may be context dependent, with further research required to elucidate the ethological implications of it during putatively positive experiences. Further research is also needed to understand the impacts of different equipment and being ridden on indicators of emotion. The findings of the current study suggest an association of licking and chewing with experience of positive, rather than negative emotion.

The literature was consistent with the findings for the Aversive preference for the following behaviours: sclera visible in cattle (Sandem et al. 2002; Proctor et al. 2015; Sandem et al. 2006), large eye aperture in sheep (Reefmann et al. 2009a) and horses (Lansade et al. 2018), ears back in sheep (Whittaker et al. 2019), and nares flared in horses (Lundblad 2018). It is likely that these behaviours are indicators of emotion in horses, with increased occurrence/rate/duration of the behaviours with a negative emotional experience.

Blinking contrasted with a study in horses (Merkies et al. 2019), but was consistent with a study in dogs (Bremhorst et al. 2019). In humans, blinking rate also increases with stress (Hall et al. 2013), which may lend support to the current experiment. Increased blink rate has also been associated with an increase in dopamine, and may reflect predisposition to stereotypy, habit formation, impulsivity, or behavioural inflexibility (McBride et al. 2017) It may be that blink rate is an indicator of emotion in horses, with an increased rate associated with a negative emotional experience.

Contraction of the underlying inner eyebrow raiser musculature leads to observable eyebrow angulation and/or wrinkles, and has been associated with pain in horses, and fear and sadness in humans (Hintze et al. 2016). The eyebrow results of the current experiment are somewhat similar to a study in horses that measured the average angle of the uppermost eyebrow wrinkle (Hintze et al. 2016). Although, the assumed valence of the treatments used in that study could not be verified due to variation within horses across the control phases, and stimulus specificity could not be ruled out. No lateral bias was found, and so observation of one eyebrow may be sufficient. Similarly, a study on horses during transport (putatively inducing negative emotional ‘stress’) found that the rate of upper eyelid raises was increased compared to control (Lundblad 2018). Eyebrow angle, and possibly eyebrow wrinkle, may be indicators of emotion in horses, with an increased rate associated with a negative emotional experience.

Most facial movements of horses are bilaterally symmetrical except for ear and nostril lift movements (Wathan et al. 2015). In the current study, the left side of horses’ heads were observed, with the ear being the only feature that was coded separately for left and right. The right ear did not behave the same way as the left with respect to forward movements. Asymmetric ear position (one ear forward, one back) was not specifically coded in the current study. However, the left ear forward and right ear back behaviours were both increased in the Aversive condition. This may imply the ears were asymmetric more often in the negative treatment, although both variables were modelled separately. Despite this, the results may support those of Lansade et al. (2018), where asymmetric ear positions were also more often observed during the negative treatment.

Horses may display auditory laterality, with horses shown to orient their right ear towards a whinny from a familiar horse and orient their left ear towards a whinny from an unknown horse (Basile et al. 2009). Although contentious, sensory and/or motor laterality of ear position is suggested to reflect cerebral lateralisation of emotional processing, with left ear movements indicating right hemisphere dominance during processing of negative emotional stimuli, and vice versa for positive emotional stimuli (Leliveld et al. 2013; Whittaker et al. 2019). In the current study, there were differences in left and right ear forward behaviours between the preference levels that align with this suggestion and the findings of (Thorbergson et al. 2016). It may be that the rate of left ear forward and back movements indicate the experience of negative emotion, and the rate of right ear forward indicates a more positive experience in horses.

Although findings of Sankey et al. (2010c) suggest that the ears forward position is associated with a positive perception of a stimulus in horses, another study found no difference between positive and negative treatments in observations of ears forward position (Lansade et al. 2018). However, increased frequency of ear position changes have been found with a negative stimulus in other species (Reefmann et al. 2009b; Reefmann et al. 2009a; Marcet-Rius et al. 2019; Whittaker et al. 2019), and ears were forward proportionately less in sheep experiencing a putatively positive grooming stimulus, instead being more passive (to the side and floppy) (Reefmann et al. 2009a). Careful interpretation across studies is required for behaviours that are measured in different ways (e.g., continuous or time sampling), with different data types (e.g., rate of movements, duration, odds of occurrence) and analysis methods.

Instead of Preference, it is possible that ear position could indicate directional attention to auditory stimuli. The position of the stimulus relative to the animal may influence ear position in the rostral-caudal (front-back) and/or left-right planes. The location of the Aversive treatment on the right side of the horse caudal to the ear might be expected to influence the results of the right ear back behaviours, rather than the left. However, the left and right ears behaved similarly for the ear back position, with an increase in frequency and duration seen with the Aversive condition compared to the Preferred. Additionally, the spray only occurred over the first 40 seconds of the 3-minute period, thus the effect of stimulus location on auditory attention over time is expected to be limited. The position of the Preferred treatments varied (none, in front, or left side), so would not be expected to consistently influence ear behaviour, lending confidence to the results.

An alternative explanation for the ear behaviour would be directional attention to extraneous noises (e.g., bird chirp, building creaking) occurring on one side of the horse, or individual differences in auditory laterality. However, noises were not systematically applied and would not be expected to bias results. The potential for the influence of individual auditory laterality (independent of emotion) is unknown. It may also be possible that animals experiencing positive emotion are less attentive in general to auditory environmental stimuli and perhaps less vigilant, particularly if relaxed (i.e., low arousal positive valence), than when experiencing negative emotion which may trigger hypervigilant behaviour. However, this would not explain the differences between left and right ear behaviour that were found.

In the current study, head shake, head tossing, left or right ear flat, face pulling, and puckered chin behaviours were not associated with treatments of differing preference or emotional experience. Support of these results is found in other research for head shake, face pulling (tongue out), and ears flat behaviours (Thorbergson et al. 2016; Lansade et al. 2018). Although, some of those behaviours did differ by individual treatment, which highlights the need to consider individual horses' regard for each treatment. It may be that these behaviours are not indicators of emotion in horses, or that the variation in the study overshadowed any true effect. Further research is suggested to corroborate the results.

9.4.2 Facial behaviour and treatment

When compared to the No-stimulus condition, several facial parameters were impacted by a particular treatment. It is interesting that 13 facial behaviours varied across all preference levels, but only one behaviour, oral investigation, varied across all treatments. The value of interpretation of results by treatment rather than by individual perception of the treatment, in this case, preference, is questionable. Nevertheless, selected behaviours are discussed below with reference to the literature.

Support may be found for the changes in oral behaviours seen with the Groom and Spray treatments (Lansade et al. 2018). The responses may be specific to Grooming and/or mimicry of mutual grooming, which may be an automatic response (Lansade et al. 2018), and are in line with contact-seeking behaviour seen when preferred areas are massaged (Feh et al. 1993; McBride et al. 2004). Oral investigatory behaviour was also increased in the Person treatment and may also be stimulus related, as some horses were observed to reach out towards the Person standing in front of them. Although, the analysis by preference, which combines results from the No-stimulus, Groom and Person treatments, is of more importance, it is good to consider the potential for stimulus specific effects. It may be that oral investigation behaviour is only associated with positive emotion in the context of the Person and Groom stimuli used in this study. However, it is reassuring that four of the five oral behaviours varied by preference, whereas only one varied by Person, and two varied by Groom treatments, and again highlights the value of analysis by preference rather than by treatment.

The findings for small eye aperture (eyes half closed) with the Groom treatment contrast to those of Lansade et al. (2018). However, there were only two treatments in that study and comparison was between a presumptively positive and negative grooming treatments. In the current study the Groom treatment was not the preferred treatment for all horses. This may explain the contrasting results and highlights the need to verify how the putatively pleasant/unpleasant stimuli are perceived by the animals, the value of having a control or intermediate condition, and in the context of the current study, the value of interpreting results based on preference rather than by individual treatments.

Some behaviours did not occur during one or more preference categories or treatments and so were unable to be analysed (e.g., yawn). A few behaviours were observed rarely, or only occurred in a small number of horses. The decision was made not to analyse some of the behaviours, such as in the case of cheek contraction (n = 8). In other cases, the behaviour was considered further in statistical models (e.g., puckered chin). When interpreting the results from events that occurred rarely, it is important to note that non-significance should not be viewed as evidence that the behaviour is not linked to emotion, as the power to detect significance would be low. Given that these behaviours tended to occur in only a small number of horses, future studies to investigate the less common behaviours should include a larger number of horses, rather than increasing the replicates in a smaller number of horses.

Camera equipment, environment, and post-processing factors can impact results. Additionally, selection of still images and sampling frame can introduce bias. In the current study, the equipment and environment (most importantly light and shadow) were analogous across recordings. No post-processing of images or selection of still images was performed, as all occurrences of behaviour were recorded from the videos. However, manual coding of facial behaviour is labour intensive and may be prone to human error, which is why intra-observer reliability analysis was performed.

9.4.3 Conclusion

In this comprehensive study of facial behaviour in horses, a surprisingly large number of facial movements varied by preference and/or treatment. The findings of the current study suggest that facial movements of horses are influenced by situations of contrasting emotional experience. More facial behaviours than body behaviours varied with preference. Support is given to the suggestion that facial movements may be more sensitive indicators of horses' subjective experience than gross body behaviours (Descovich et al. 2017; Lansade et al. 2018).

In horses, movements of the eyebrow, nares, lips, and chin, position of ears, eye aperture, blink, and sclera, and other oral behaviours were found to differ with preference for a stimulus. These facial behaviours may be indicators of emotion, at least in the context of this study. A more positive emotion may be indicated by increased counts of right ear forward, left or right ear to the side, small or neutral eye aperture, licking and chewing, or oral investigation, durations of left or right ear forward or to the side, or probability of a drooping lower lip behaviours. A more negative emotion may be indicated by increased counts of left ear forward, left or right ear back, blink, sclera visible, large eye aperture, eyebrow angled, eyebrow wrinkle, or nares flared, durations of left or right ear back, or nares flared, or probability of contracted lips behaviours.

The findings support the notion that facial behaviours can be used in characterisation of animal emotional experience, at least in species that have facial musculature that enables distinguishable facial movements. Although the influence of the nature of the stimulus must also be considered for how it may encourage or limit certain behaviours (stimulus specificity), in horses, facial movements may be indicators of emotion, suggesting the existence of so called facial expresssions in horses.

At the time the study was conducted, many of the facial behaviours had not been previously studied. To the author’s knowledge, this study presents the most comprehensive investigation of facial behaviour with regard to positive and negative emotional experience to date. The findings may also provide objective support to those of other studies on animal behaviour that may not have independently identified emotional valence and/or included three levels of valence. Future research may involve further use of remote monitored devices and/or artificial intelligence to remove any potential impact of human influence and improve efficiency of data collection and interpretation.

Chapter 10 Interpretation of heart rate variability, eye temperature, body behaviour, and facial movements as indicators of emotional valence and arousal in horses

10.1 Introduction

The aim of the experiment presented in the previous Chapters 6-9, was to investigate whether and how physiology, body behaviour, and facial behaviour parameters varied in horses undergoing a pleasant experience, compared to a mildly aversive, and an intermediate experience. The outcome was to determine if the parameters could be non-invasive indicators of emotional experience.

This chapter combines and extends the results from Chapters 7-9 and attempts to further characterise the indicators of emotional experience in terms of the arousal and valence dimensions of emotion. An overview of the combined results is given initially to enable assimilation of the many significant parameters along with the relative impact size. An interpretation of the parameters that differed across all three preference levels is given. The chapter concludes with discussion of the findings and potential limitations of the experiment.

When attempting to identify emotional valence, the potential effects of emotional arousal may also be considered (Figure 1 and Figure 2). In the current experiment, preference for a treatment was used to identify the likely relative valence level on an individual basis (Chapter 6). It might be expected that parameters reflecting valence would change sequentially from the least to the most positive valence level, i.e., Aversive to Intermediate to Preferred levels of preference.

Positive emotional stimuli are less salient than negative ones (Smith et al. 2016). Arousal or attention level may be highest to the most salient stimulus and lowest for the least meaningful one. In this experiment, it is suggested that the intermediate preference was likely to induce the lowest arousal level (least salient), the aversive to induce the highest arousal level (most salient), and the preferred treatment an arousal level in-between (less salient than the aversive but more that the intermediate treatments). This suggestion is based on ethology, the theorised relative salience and survival impact, and the results of other studies suggesting that positive emotion inducing stimuli such as stroking/grooming (Janczarek et al. 2018; Lansade et al. 2018), or the presence of a familiar human, are emotionally arousing (Tamioso et al. 2018). It might be expected that parameters reflecting arousal would change sequentially from most to least salient, i.e., Aversive, to Preferred, to Intermediate preference level.

The aim for this chapter was to further characterise indicators of emotional experience in horses in terms of the arousal and valence dimensions of emotion.

10.2 Overview of results

The significant results from Chapters 7-9 are presented in a simplified overview in Table 28, which shows the model predicted mean value and the percentage change from the Preferred treatment. In total, 62 parameters were measured, 48 were analysed, and 45 parameters differed with preference. Compared to the Preferred treatment, the Aversive treatment was associated with alteration to more parameters (n = 44) than the Intermediately preferred treatments (n = 17). The Aversive treatment resulted in larger impact than the Intermediately preferred treatments.

The percentage change in predicted values across the parameters ranged widely. Although statistically significant, small changes may or may not be of clinical or practical significance, for example, a -0.3% change in average eye temperature between the preferred and aversive conditions.

10.2.1 Most preferred versus aversive treatment

When compared to the most Preferred treatment, the least preferred (Aversive) treatment resulted in a higher average heart rate, overall heart rate variability (SDNN), lower RMSSD:SDNN ratio, and average eye temperature, suggesting a shift in the balance towards the sympathetic nervous system. There were increases in leg movements, tail swishing, tail thrashing, and extremes of neck movement, with longer time above neutral and less time below neutral. It was associated with the left ear moving forward and both ears moving back more frequently, while the rate of right ear forward was decreased. Interestingly, both ears were forward for less proportion of the time, were back for longer, and were to the side less frequently, for less time. For movements involving the eye, there was a higher probability of observing a large eye aperture, and a lower frequency of small eye aperture. Blinking, an angled eyebrow, and a wrinkled eyebrow were more frequent, and the sclera (eye white) was also visible more frequently with the Aversive preference. While there was more chance of contracted (pursed) lips occurring, there was less chance of other oral behaviours occurring (drooping lower lip, oral investigation), less licking and chewing, and a lower frequency of chin wobbles. Horses experiencing the Aversive stimulus had their nares flared more often, for a longer time.

Conversely, a horse experiencing a non-aversive (Preferred or Intermediate treatment) stimulus may have a lower heart rate, less hindleg and foreleg movements, their neck below neutral for a longer time, above neutral for less time, their ears to the side more frequently, for a longer time, ears forward for longer, a small eye aperture, and show more movements of the mouth (drooping lower lip, lip licking and chewing, and oral investigation), and/or chin wobble behaviours. A tail swish or thrash, ears back, blink, large eye aperture, visible sclera, a wrinkled or angled eyebrow, and flared nares may be displayed less often.

10.2.2 Most preferred versus intermediate preference

When compared to the most Preferred treatment, the Intermediately preferred treatments resulted in more ear forward and ear back movements, flared or neutral nares, angled eyebrow, and blinking, and/or a decrease in neck very high movements, small eye aperture, oral investigation, chin wobbles, and lower odds of contracted lips occurring. There were also tendencies towards a decrease in the RMSSD measure of heart rate variability and tail swishing frequency, and an increase in neck very low and right ear side movements.

Table 28 Overview of results from Chapters 7-9 for the effect of preference for a treatment showing model predicted values and percentage change for statistically significant comparisons.
Analysis was based on comparison to the Preferred treatment (referent).

Category	Variable	Preferred (REF)	Intermediate		Aversive	
		Predicted	Predicted	Change %	Predicted	Change %
Heart	Average heart rate (HR) bpm	39.7	-		49.9	26
	Average interval between heart beats (IBI) ms	1576	-		1292	-18
	Overall variance in IBI (SDNN) ms	52.0	-		60.0	15
	Short term variance in IBI (RMSSD) ms	56.4	52.5^	-7	-	
	Ratio of short term to overall variation in IBI (RMSSD:SDNN)	1.1	-		1.0	-13
Eye Temp	Average eye temperature (ET) at 15oC air temp oC	36.1	-		36.0	-0.3
Leg	Foreleg lift count/3min	3.1	-		6.9	123
	Hindleg lift count/3min	2.1	-		7.7	267
	Hindleg kick probability %	0.6	-		4.0	567
	Hindleg rest count/3min	2.0	-		3.9	95
Neck	Very low count/3min	2.4	2.8^	17	3.4	42
	Low time %	11.1	-		6.3	-43
	Neutral time %	59.7	-		46.0	-23
	High time %	24.6	-		43.6	77
	Very high count/3min	0.2	0.1	-50	0.3	50
Tail	Tail swish count/3min	0.8	0.6^	-25	2.9	263
	Tail thrash probability %	4.1	-		14.7	259
Ears	Left ear forward count/3min	12.7	14.4	13	15.6	23
	Right ear forward count/3min	14.2	15.6	10	12.3	-13
	Left ear forward time %	25.4	-		19.8	-24
	Right ear forward time %	23.0	-		11.8	-49

Category	Variable	Preferred (REF)	Intermediate		Aversive	
		Predicted	Predicted	Change %	Predicted	Change %
	Left ear side count/3min	17.9	-		16.3^	-9
	Right ear side count/3min	17.7	18.9^	7	15.6	-12
	Left ear side time %	36.7	-		18.6	-49
	Right ear side time %	36.6	-		20.4	-44
	Left ear back count/3min	9.9	11.0	11	17.6	78
	Right ear back count/3min	10.5	12.7	21	18.1	72
	Left ear back time %	35.4	-		61.1	73
	Right ear back time %	38.2	-		67.2	76
Eye	Blink count/3min	41.1	43.8	7	48.1	17
	Sclera visible count/3min	3.9	-		5.1	31
	Small eye aperture count/3min	0.9	0.5	-44	0.4	-56
	Neutral eye aperture count/3min	6.8	-		9.6	41
	Large eye aperture probability %	1.7	-		32.6	1819
	Eyebrow wrinkle count/3min	1.7	-		2.2	29
	Eyebrow angled count/3min	1.4	1.8	29	2.7	93
Mouth	Contracted lips probability %	8.3	2.8	-66	51.5	521
	Drooping lower lip probability %	33.1	-		8.0	-76
	Licking or chewing count/3min	4.0	-		2.1	-48
	Oral investigation count/3min	1.2	0.8	-33	0.3	-75
Chin	Chin wobble count/3min	4.0	3.4	-15	2.3	-43
Nares	Nares flared count/3min	6.3	7.0	11	14.2	125
	Nares flared time %	4.3	-		7.0	63
	Nares neutral count/3min	14.0	15.1^	8	27.6	97

^ $p \leq 0.10$, all others were significant at $p \leq 0.05$
- No significant findings

10.3 Interpretation of results by arousal and valence dimensions

There were 16 parameters that differed across all three preference categories (Table 29). They were all behavioural parameters and included three body behaviours and 13 facial behaviours. Parameters that varied sequentially from the Intermediate, to the Preferred, to the Aversive treatment, may indicate arousal. Parameters that varied sequentially from the Preferred, to the Intermediate, to the Aversive treatment, may indicate valence. However, there were only tendencies towards significance in the Intermediate versus Preferred comparisons for the neck very low, tail swish, right ear forward, and nares neutral rates.

Table 29 Table of 16 parameters that varied across all three preference categories showing the direction of change and interpretation as indicators of arousal or valence dimensions of emotion.
All parameters are rates of behaviour (count/3 minutes) except contracted lips which is a probability of occurrence (%). Figures are from model predicted values. The direction of the effect is shown (“<” signifies an increase with arousal levels or from negative to positive valence, whereas “>” signifies a decrease)

Dimension	Parameter	Lowest (Intermediate)	Mid (Preferred)	Highest (Aversive)	Direction of change
Arousal	Neck very high	0.1	0.2	0.3	<
	Tail swishing	0.6*	0.8	2.9	<
	Contracted lips probability	2.8*	8.3	51.5	<
	Right ear forward	15.6	14.2	12.3	>
	Right ear side	18.9*	17.7	15.6	>
		Negative (Aversive)	**Mid (Intermediate)**	**Positive (Preferred)**	
Valence	Small eye aperture	0.4	0.5	0.9	<
	Chin wobble	2.3	3.4	4.0	<
	Oral investigation	0.3	0.8	1.2	<
	Neck very low	3.4	2.8*	2.4	>
	Left ear forward	15.6	14.4	12.7	>
	Left ear back	17.6	11.0	9.9	>
	Right ear back	18.1	12.7	10.5	>
	Blink	48.1	43.8	41.1	>
	Angled eyebrow	2.7	1.8	1.4	>
	Nares flared	14.2	7.0*	6.3	>
	Nares neutral	27.6	15.1*	14.0	>

* Denotes a p value of ≤ 0.10

Indicators of emotional valence

In the current study, the frequency of left ear forward, left ear back, right ear back, blinking, angled eyebrow, nares flared, and nares neutral behaviours increased from the Preferred, to the Intermediate, to the Aversive treatment. These behaviours may be indicative of emotional valence, with a decrease from negative to positive valence. The rate of chin wobbles, small eye aperture, and oral investigation behaviour may indicate emotional valence, with an increase from negative to positive valence.

Indicators of arousal

In the context of this study, indicators of emotional arousal may be rates of neck very high, tail swishing, and the probability of a contracted mouth, which all increased with increasing arousal. The rates of right ear forward and right ear side behaviour may also indicate arousal and show a decrease with increasing arousal.

10.4 Discussion

This experiment found that many behavioural and some physiological parameters varied with differing emotional experiences. The observed changes were catalogued and discussed in terms of positive and negative emotional experience. As expected, it was easier to differentiate large contrasts of emotion such as that between positive and negative, but harder to differentiate positive and intermediate emotional experience. In the future, it may be that refinements in measurement of responses, perhaps using automation of behaviour measurement and machine learning techniques, allow for finer granularity in these more subtle changes.

It was hypothesised that the Preferred treatment would increase parasympathetic nervous system activity, assessed using eye temperature and heart rate variability (Boissy et al. 2007b), and induce behaviours associated with a relaxed state (low arousal). However, horses experiencing a Preferred stimulus did not display behaviours associated with low arousal. Some of the behaviours displayed may indicate increased arousal, which is supported by findings from recent studies. Although the heart rate variability (RMSSD) results showed a tendency suggesting more parasympathetic activity with the Preferred treatment, and there was increased sympathetic activity with the Aversive treatment, insufficient evidence was found in this study to support the hypothesis. Other studies have found significant behavioural differences in the absence of significant physiological differences in ridden horses (Thorbergson et al. 2016). However, further investigation may be warranted and may include investigation of the duration of physiological responses to emotional experiences.

Eye temperature did not respond as expected, with no impact of preference on maximum temperature, and only a small decrease in average temperature seen with the Aversive treatment. Despite the unexpected results, the study serves as basis for further investigation of whether eye temperature could be a marker of positive and/or negative emotion in horses.

The forty-four behavioural and physiological parameters that varied by preference and the underlying emotional experiences, were discussed in the relevant chapters (Chapter 7-9). Further interpretation of the results with reference to the arousal and valence dimensions of emotion was based on theorised relative arousal levels of the preference levels. It was argued that the Aversive stimulus was likely to induce a higher level of arousal, and the Intermediate a lower level of arousal, than the Preferred stimulus due to differences in salience of the stimuli for each horse. Of the 16 parameters that varied across all three preference levels, five were determined to be indicators of emotional arousal: rates of neck very high, tail swishing, right ear forward, and right ear side, and probability of contracted lips behaviours. The remaining eleven parameters were determined to be indicators of emotional valence: the rates of neck very low, small eye aperture, blinking, angled eyebrow, nares flared, nares neutral, left ear forward, left ear back, right ear back, chin wobble, and oral investigation behaviours. Further research is required to examine if the apparent tendency towards differences between the Preferred and Intermediate stimuli for four of the parameters hold true. These findings are a foundation for further research.

No other studies have used three levels of independently verified valence with a mechanism for interpretation of responses with respect to the arousal and valence dimensions of emotion in horses, and investigated body and facial behaviours in depth, to the authors knowledge. However, mention must be made of a study focusing on horse whinnies wherein Briefer et al. (2015b) concluded that head height was an indicator of valence rather than arousal. The levels of valence were presumed. Although they similarly found that in the putatively negative emotionally valenced situation (separation from conspecifics), horse's heads were held in the high position for longer than during the putatively positive situation (reunion with conspecific, i.e., relief from separation), when an arguable attempt was made to control for arousal level, it appeared that this was due to emotional valence. It may be that neck height as an indicator of valence is arousal level and/or context dependent; or alternatively, there may have been a stretch in interpretation due to flaws in the method used to control for arousal, or in fitting the arousal or valence model. It also noted that horses engaged in locomotory activity, and the putatively positive treatment (reuniting) involved correction of a negative situation (separation), whereas in the current study horses where stationary, and the positive treatment was not contingent on prior application of a negative one. These differences may also explain the differences in findings. Neck/head position is related to attention to the environment, enhancing auditory, visual, and olfactory sensation (McGreevy 2004). The neck may be expected to be in the higher positions more often or for longer when scanning for threats, or when attentive to a stimulus, whether positively or negatively valenced (i.e., higher arousal), and less often when there is no threat, or less attention towards a stimulus (i.e., lower arousal). Which lends support to the findings of the current study that rate of neck very high behaviour is positively related to arousal level.

There may be interplay between arousal level and emotional valence, which is difficult to separate (Kremer et al. 2020). For example, it is suggested that, compared to the intermediate treatment, animals experiencing positive emotion associated with a Preferred treatment in this study were more attentive to the stimulus (higher arousal) than the Intermediate, but less attentive to auditory environmental stimuli, and perhaps less vigilant, than with the Aversive stimulus. It may be that positive emotion decreases threat concern. Further research is required to confirm and explore interaction between arousal and valence.

Emotion may impact different measures of the same behaviour in different ways. For example, the Aversive treatment was also associated with the highest time spent with neck above neutral, the lowest time spent at neutral, and the lowest time spent below neutral, but the highest rates of neck very low and neck very high behaviour. It could be that frequencies and durations of behaviour provide different information about the arousal and valence dimensions of emotion. Analysis of the different data types: frequency and proportion of time, for the same behaviour may provide a richer understanding of the impact of preference/emotion than either alone.

This research did not set out to determine what most horses do or do not enjoy. This research was designed to examine a subset of horses that cared about the offered stimuli and record their behavioural and physiological responses. The treatments differed in many aspects, including the proximity and/or involvement of a human, and degree of control the horse had over the treatment (voluntary/involuntary), as is the case with other studies (Lansade et al.

2018). It is acknowledged that horses' individual preferences for the treatments may have been related to these aspects.

Alternative explanations

It is important to consider alternative explanations for the results and the generalisability of the findings. Instead of being indicators of affective experience, emotional valence, and/or arousal, it may be that the responses seen in this study were stimulus specific (e.g., neck extension above neutral with grooming), species specific (e.g., tail swish), and/or reflect a display of behavioural intention (e.g., escape from threat via fight or flight). Further research is required to corroborate and expand on the results.

It has been suggested that tactile stimulation, such as massage or grooming, may trigger an unconscious autonomic effect on cardiac variables, as the stellate ganglion, which has cardiac efferents, is located close to the cranial wither area (a preferred grooming site in horses) (Feh et al. 1993). This could cause a stimulus specific effect of a decrease in heart rate, independent of emotion, but no such decrease was evident in this study or others (Janczarek et al. 2018b; Lansade et al. 2018).

The nature of the emotion inducing stimulus may limit or encourage certain behaviours. For example, horses were observed to extend their head and neck upwards during the Groom treatment, and there was an increase in the neck very high behaviour, and the likelihood of lip licking and chewing, perhaps in an attempt at mutual grooming (McGreevy 2004). It is possible that the behavioural changes are specific to the type of stimulus, and thus are not universally applicable (Lansade et al. 2018). Results should be interpreted by individual preference or underlying emotional regard. The inclusion of several types of stimuli in the current study reduced potential impact of stimulus specificity. However, further research is encouraged to determine if the findings of this study are particular to the stimuli used, or if they may be generalised to other aversive or pleasant stimuli.

The startle response is a reliable indicator of emotional valence in humans (Mauss et al. 2009). However, fear or anxiety caused by a sudden, novel, and unpredictable stimulus, in this case the aerosol spray, might cause a different behavioural reaction than another type of aversive stimulus, for example, pain from an injection. Future research could use positive, intermediate, and negative emotion inducing stimuli that use the same substrate/stimulus type, with constant, rather than intermittent, application and response.

Habituation to an aversive stimulus can impact results where there is repeated exposure. The horses were purposefully only exposed to the aversive stimulus for three spray bursts on three occasions during this experiment. Although the horses were also exposed to the stimulus during the previous experiment, with the limited exposures and short, sudden, and unpredictable nature of the stimulus, it is considered unlikely that horses would have habituated. The heart rate results showed a large response to the aversive treatment, which supports this assertion. Observation of the horse’s behavioural response also suggested that no appreciable habituation occurred.

The duration of treatment may impact momentary emotion and/or emotional state, and therefore influence results (Hintze et al. 2016). The duration of stimulus application should be similar across treatments. However, application of a prolonged aversive stimulus may not be possible due to ethical and safety considerations, and application of a very short duration positive stimulus may not develop sufficient response (Thorbergson et al. 2016). Additionally, the premature removal of a short positive stimulus may lead to negative emotion, as in frustration with food removal after a short time. The treatment duration for the current study was based on the literature (which ranged from 30 seconds to 10 minutes) and balancing the aim of detecting changes associated with momentary emotion, rather than emotional state, with limiting exposure to the negative condition. Factors such as sex, age, emotional excitability, size (horse/pony), or breed, may also have an effect which warrants further investigation (Thorbergson et al. 2016; Janczarek et al. 2018b). Careful consideration of the choices offered as emotion inducing stimuli is required and should take into account individual preferences, as well as relative levels of arousal and valence, with duration of application and timing of measurements.

To what extent the results of this study are generalisable to horses beyond those that made a reliable choice during preference testing, should also be considered. Whilst further research is required to replicate the findings and confirm the tendencies seen in this study, the results provide a promising foundation and an approach to distinguishing indicators of valence from indicators of arousal based on preference.

10.4.1 Limitations

Establishing the validity of behavioural indicators of positive emotional valence is difficult, as there are no gold standard markers with which to correlate and establish sensitivity and specificity (Whittaker et al. 2019). In this developing field of research, it is important to consider the potential limitations of this study. The research approach for this thesis, which includes this experiment and the underpinning assumptions, is considered further in Chapter 12.1.

Attempts were made to control potential confounding factors, such as effects of the experimental environment, personnel, and equipment, through habituation and training sessions, using experimenters previously unfamiliar to the horses, and taking care with horse interactions. Movement and exercise, which can impact cardiac variables, were limited by restraint in the stocks. Time of day and treatment application were consistent across replicates, and social isolation was avoided. The Spray treatment may have had a carryover effect on subsequent treatments during the same replicate, thus the order of treatment was randomised for Groom and Person, and the Groom, Person and No-stimulus treatment periods used for analysis were selected at random. However, there were experimenters nearby during all treatments and human influence through factors such as interspecific emotional contagion via chemosignals cannot be ruled out (Semin et al. 2019; Calvi et al. 2020). In future, it may be possible that technological advances in equipment and biosensors allow for conduct of similar experiments without the influence and potential confounding effects of humans (Neethirajan et al. 2021).

An attempt was initially made to include a control treatment (the No-stimulus treatment) as a reference point for the changes seen with the other treatments. A control treatment would ideally be of neutral emotional valence and have a consistent, known arousal level, that is, it would not elicit a positive or a negative emotion, and would elicit the same response from all horses. However, a true control treatment is not possible as, for example, a No-stimulus treatment may not elicit the same type and strength of emotion across all animals, because animals' emotional responses are specific to the individual. There is a continuum between the most positive and the most negative experience for every horse. The problem is that a more neutral (less positive or less negative) experience for one horse is not the equivalent for another horse. Thus, there is no universal control for comparison between horses. Until a scale is developed for rating how positive (or negative) an experience is for an animal, then it is difficult to directly compare between animals.

There are no gold standard emotion inducing stimuli for this field of research which can be used or compared to. Additionally, in this study, the No-stimulus treatment was the preferred treatment for some of the horses. Future research could include a no-stimulus treatment and exclude horses that preferred this treatment over other putatively pleasant experiences. However, future research should still recognise that a universal control situation does not currently exist.

In this study, two treatments were labelled as Intermediate preference for each individual horse because they were not the most preferred and were not as aversive as the spray. The two treatments may be perceived differently by each horse, which may have increased the variation and account for why there were fewer differences found between the Intermediate and the Preferred treatments. It could also be argued that none of the stimuli were absolutely positive, and that the most preferred treatment was just the least negative. The use of naturally occurring episodes of presumed positive emotion, such as voluntary exploration, grazing, automated grooming brushes, or play, maybe an alternative approach to experimentally inducing emotion (Guesgen et al. 2017).

Although statistically significant, the low predicted counts for some parameters might limit their utility in field applications. Differentiation using these parameters on an individual horse basis may be difficult over the short time span associated with momentary emotion (as distinct from emotional state).

Limitations of analysis

The behavioural and physiological responses of horses varied by treatment and preference. There may be an effect of treatment within preference. However, due to collinearity between the Spray and Aversive treatments, it was not possible to analyse this. Repetition of the study using only horses that displayed the same treatment preferences may help to address this issue.

The number of observations differed across treatment and preference categories. The Intermediate category included the two treatments that were neither the aversive nor most often preferred, whereas the Preferred and Aversive categories were comprised of one treatment each. Two occurrences of the No-stimulus, Groom, and Person treatments, and one Spray treatment, were analysed per replicate. There is a possibility that there may have been an effect on the statistical models from an unbalanced number of observations.

Due to the novelty of the research approach and dearth of comparable published literature at the time, conducting meaningful *a priori* power analysis was difficult. Some behaviours occurred infrequently or in a small number of horses. Future research may benefit from increasing the number of subjects, rather than the number of replicates.

The intra-observer reliability of behavioural observations was extremely high (Martin et al. 2007; Petrie et al. 2013). However, as is the case in other studies, the observer could not be blinded to the treatments while coding behaviour, and that cannot be excluded as a source of bias, even though most of the behaviours measured were clearly distinguishable (Lansade et al. 2018). Due to the more subtle nature of some of the behaviours, and the potential for observer bias, future studies may benefit from also assessing the inter-observer reliability of the behaviours (Thorbergson et al. 2016).

The results of this study are relative findings, not absolute. Although model predicted values were given, due to the significant effect of the horse in the models, they should not be relied on as absolute values. Emotional arousal and/or valence can only be inferred by comparison of changes in behaviour within a subject. Thus, based on the results of this study, it is possible to observe a horse and determine if they are experiencing a Preferred (positive valence), Intermediate (intermediate valence), or Aversive (negative valence) stimulus when they are observed experiencing stimuli of differing valence and/or arousal levels.

10.4.2 Conclusion

This study used a novel approach to investigate how physiological and behavioural responses of horses changed whilst undergoing experiences, inducing positive, negative, and intermediately valenced emotion. The negative emotional condition resulted in changes to many behavioural and physiological measures. Fewer differences were seen between the intermediate and positive conditions. The hypothesis that the positive emotional condition would increase parasympathetic nervous system activity, or induce behaviours associated with a relaxed state, was not supported.

Within the study context, it was possible to differentiate between a horse experiencing an aversive stimulus, and one experiencing a non-aversive stimulus, based on physiological and behavioural responses. The results suggest that the following may be useful indicators of contrasting affective experience in horses: heart rate, heart rate variability, eye temperature, and behaviours involving legs, neck, tail, ears, eyes, eyebrows, mouth, chin, and nares. Whether changes in these variables indicate differences in arousal or emotional valence needs to be considered, along with stimulus specific effects. It may be that the rate of a behaviour and its duration indicate different things.

Sixteen behavioural variables differed across all three preference categories, although only trends were established between the Preferred and Intermediate condition for four of those variables. Information about arousal may be provided by observation of the frequency of neck movements to the very high position, tail swishing behaviour, and the odds of a contracted mouth, with rates increasing with arousal level, and right ear forward and side behaviour, with rates decreasing with increased arousal. Emotional valence may be indicated by the frequency of neck very low, left ear forward, left and right ear back, blinking, angled eyebrow, nares flared, and nares neutral behaviours, with rates decreasing from negative to positive valence, and chin wobble, small eye aperture, and oral investigation behaviour, with rates increasing from negative to positive valence. Further research is required to corroborate the trends seen and to investigate the generalisability of the results beyond the stimuli used in this study.

Facial movements were examined in detail and found to vary with emotional valence and/or arousal. Facial movements are expressive of emotion in horses, that is, horses display facial expressions of emotion. In line with other studies, this work expands on existing knowledge and confirms that research on facial movement or expressions may be useful for identifying animal emotions.

Finally, this study explored an emerging area of animal welfare science. It highlighted some of the challenges of research on indicators of emotion, such as the use of sufficiently powerful pleasant and aversive treatments that are ethically acceptable and practical in the experimental setting, the non-existence of a universal emotional neutral point, and difficulties untangling the effects of emotional valence and arousal without gold standards to compare to.

Chapter 11 Multivariate Analysis

11.1 Introduction

In Chapter 10 several parameters were suggested to be indicators of the arousal or valence dimension of emotion in horses via interpretation of the findings from Chapters 7-9. One of the limitations of the univariable analyses that were conducted in Chapters 7-9 is that comparisons between parameters, or patterns of response, are unable to be established. This chapter builds on Chapter 10 by using multivariate analyses to summarise the data, explore relationships between individual parameters, and/or be used to refine future research or application to a smaller number of distinct factors or parameters.

The aim was to examine the data for clustering using multivariate analysis methods, describe the clustering, determine if the clusters differed by the explanatory variable (preference), and identify a smaller set of parameters that may be observed in future as indices for the clusters.

11.2 Materials and methods

The data set was comprised of the count data from the 16 parameters that were found to vary across the three preference categories in Chapters 7 to 10. Note that the contracted lips behaviour was analysed/reported in Chapters 9 and 10 as the probability of occurrence, but due to the methods of multivariate analysis performed in the current chapter, the count data for this behaviour was used. The per minute rates of behaviour were calculated from the count data accounting for the time that the body part was out of view. Three data sets were subsequently examined: the first one contained all 16 significant parameters, the second one contained the five parameters that were found during univariable analysis to indicate arousal, and the third one contained the remaining 11 parameters which were found to indicate valence (Table 30).

Table 30 Indicators of emotional arousal or valence as determined in Chapter 10.

Dimension	Parameter	Relationship with dimension
Arousal	Neck very high	Positive
	Tail swish	Positive
	Contracted lips*	Positive
	Right ear forward	Inverse
	Right ear side	Inverse
Valence	Chin wobble	Positive
	Oral investigation	Positive
	Small eye aperture	Positive
	Left ear forward	Inverse
	Left ear back	Inverse
	Right ear back	Inverse
	Nares flared	Inverse
	Nares neutral	Inverse
	Neck very low	Inverse
	Blink	Inverse
	Angled eyebrow	Inverse

* Contracted lips behaviour was significant for probability of occurrence; the other behaviours were counts/3min.

11.2.1 Data analysis

Exploratory factor analysis, conducted using principal component factor (PCF) and principal component analysis (PCA), was used to identify and then predict the underlying component structure of the parameters. The resulting structure following PCF identifies the cause-relationship between the latent variables and the observed ones, with the structure from PCA representing the linear combination of variables. Both the PCF and PCA represent the new factors or components, respectively, as eigenvalues. All factors or components with eigenvalues >1.0 were retained, the scree plot was examined for the point of inflexion, and components or factors to the left of the point of inflexion were retained for further analysis. Finally, the Kaiser-Mayer-Olkin (KMO) index was generated, and all parameters over >0.5

retained. Otherwise, the parameter was removed and the PCF or PCA rerun, with eigenvalues checked and scree plots generated until all parameters had a KMO index >0.5.

For the PCF only, extracted factors were obliquely rotated to simplify the factor structure. Following rotation, parameters were retained within each factor if the factor loading was not between -0.4 and 0.4 ($|r| > 0.40$), unless otherwise noted.

For the PCA only, parameters were retained within each factor if the factor loading was not between -0.2 and 0.2 ($|r| > 0.20$), unless otherwise noted. Predicted component scores were calculated for each observation using the regression method, which was based on the component loadings and the response for the constituent parameters within each observation (Field et al., 2012). The score is a numerical value calculated from combination of the rate data and loading coefficient of each parameter for a particular component.

For both PCA and PCF, the final components or factors, respectively, were named/labelled to indicate the theme of the component/factor to aid in interpretation. PCA and PCF analyses were conducted in Stata Statistical Software (Release 17, StataCorp LLC, College Station, Texas, USA).

A mixed-effects model (linear regression) was used to examine the relationship between preference and each PCA score (Zuur et al. 2009). The model was necessary to account for the impact of preference, while accounting for the repeat measurements on horses on different days, giving conservative estimates of the standard errors (SE). As the model accounts for repeated measures and repeated analyses did not occur, no further correction to SE or p value was necessary.

This analysis was performed in R: A Language and Environment for Statistical Computing (version 4.1.1, R Core Team, The R Foundation for Statistical Computing, Vienna, Austria) with horse, and replicate within horse as random effects to account for the baseline in each session. In the final models, the assumptions of normality and homoscedasticity of the errors were assessed by graphical analysis of the residuals. Results from the Preferred category were compared to those from the Intermediate category and to those from the Aversive category. The significance of relationships was assessed using the log likelihood ratio test and Wald test statistic. Statistical significance was set at $p \leq 0.05$.

11.3 Results

11.3.1 Principal component factor analysis

Analysis of the 16 parameters that varied across the three preference levels (PCF1)

The results of PCF1, rotated analysis of the 16 parameters that were found to vary across all levels of preference, are presented in Table 31. One parameter, angled eyebrow, was removed following rotation as the coefficient values no longer met the threshold for inclusion in any of the factors. Five factors were included in the model that collectively explain 67.4% of the variation in the data. The factors have been named based on the major contributors to each factor. The results of the unrotated PCF analysis are presented for completeness in Appendix A.7 Table A 40.

Table 31 PCF1 - Results from exploratory factor analysis using PCF (rotated) on the parameters that varied across the three preference levels during univariable analyses, showing correlation coefficients within each factor.
All parameters are per minute rates of behaviour. Collectively, the five factors retained in the model explain 67.4% of the variation in the data. Note that angled eyebrow, was removed following rotation as the coefficient values no longer met the threshold for inclusion in any of the factors.

Factor Name	Parameter (rates)	Emotion indicator*	Factor 1 17.6%	Factor 2 16.6%	Factor 3 14.0%	Factor 4 10.0%	Factor 5 9.2%
Ears forward-side	Right ear forward	AI	0.9005				
	Left ear forward	VI	0.8244				
	Right ear side	AI	0.7833				
Ears back-nares	Left ear back	VI		0.9449			
	Right ear back	VI		0.9105			
	Nares flared	VI		0.6114			
	Nares neutral	VI		0.5035			
Neck-tail	Neck very high	AP			0.8300		
	Tail swish	AP			0.7268		
	Neck very low	VI			0.6587		0.4473
Blink-chin	Blink	VI				0.7262	
	Contracted lips	AP				0.6232	
	Chin wobble	VP				-0.6824	
Eye-Oral	Oral investigation	VP					0.824
	Small eye aperture	VP					0.5727

*Parameters were determined in Chapter 10 to indicate arousal (A) or valence (V) from univariate regression analysis. A = indicator of arousal, V = indicator of valence, P = positive indicator, i.e., increases with increasing Arousal/Valence, I = inverse indicator, i.e., decreases with increasing Arousal/Valence.

Analysis of indicators of valence (PCF2)

The results of PCF2, rotated analysis on the indicators of Valence, are presented in Table 32. There were four factors retained in the model, which collectively explain 67.0% of the variance. Factors 2 and 3 overlap slightly, with angled eyebrow contributing weakly to both factors. Factors 3 and 4 overlap with chin wobble contributing weakly to Factor 3 and more strongly to Factor 4.

Factor 1 (21.1% of variance), named ears back, Factor 2 (18.8% of variance), named eye-nares, and Factor 4 (11.1% of variance) may have an inverse relationship with valence level. Factor 3 (16.1% of variance) named mouth-neck, also may have an inverse relationship with valence level. However, one parameter, oral investigation, which also has the largest loading value for Factor 3, on its own was found to have a positive relationship with valence level. It is not known why there is an apparent conflict in results for oral investigation.

Table 32 PCF2 - Results from exploratory factor analysis using PCF (rotated) on the frequency (rate per minute) of parameters that during univariable regression analyses were determined to indicate emotional Valence.
Collectively, the four factors explain 67.0% of the variation in the data.

Factor Name	Parameter	Emotion indicator*	Factor 1 21.0%	Factor 2 18.8%	Factor 3 16.1%	Factor 4 11.1%
Ears back	Right ear back	VI	0.9513			
	Left ear back	VI	0.9478			
	Nares flared	VI	0.4652	0.6813		
Eye-nares	Small eye aperture	VP		-0.7201		
	Nares neutral	VI		0.6892		
	Left ear forward	VI		0.5276		
	Eyebrow angled	VI		0.4626	0.397	
Mouth-neck	Oral investigation	VP			0.8049	
	Neck very low	VI			0.7514	
	Chin wobble	VP			-0.4889	-0.6322
Blink/chin	Blink	VI				0.8783

*Parameters were determined in Chapter 10 to indicate valence (V) from univariable regression analysis. V = indicator of valence, P = positive indicator, i.e., increases with increasing Valence, I = inverse indicator, i.e., decreases with increasing Valence.

Analysis of indicators of arousal (PCF3)

The results of PCF3, rotated analysis on the indicators of Arousal, are presented in Table 33 There were two factors retained in the model, which collectively explain 65.1% of the variance. Factors 1 and 2 overlap slightly, with contracted lips appearing weakly in both factors. Factor 1 (34.5% of variance) was named ears and may have an inverse relationship with arousal level. Factor 2 (30.6% of variance) was named tail-neck and may have a positive relationship with arousal level.

Table 33 PCF3 - Results from exploratory factor analysis using PCF (rotated) on the frequency (rate per minute) of parameters that during univariable regression analyses were determined to indicate emotional Arousal.
Collectively, the two factors explain 65.1% of the variation in the data.

Factor Name	Parameter	Emotion indicator*	Factor 1 34.5%	Factor 2 30.6%
Ears	Right ear forward	AI	0.8772	
	Right ear side	AI	0.8469	
	Contracted lips	AP	-0.3975	0.4000
Tail-neck	Tail swish	AP		0.8481
	Neck very high	AP		0.7808

*Parameters were determined to indicate arousal (A) from univariable regression analysis. A = indicator of arousal, P = positive indicator, i.e., increases with increasing Arousal, I = inverse indicator, i.e., decreases with increasing Arousal.

11.3.2 Principal component analysis score and loading plots

The results of principal component analysis (PCA) of the 16 parameters varied across the three preference levels during univariable analysis are presented in Appendix A.7 Table A 41. Five components were retained in the model, which collectively explain 70.5% of the variation in the data. One parameter, blink, was removed as the values did not meet the threshold for inclusion in any of the factors.

PCA score plots of preference from the largest two principal components (PC1 and PC2) for each analysis were examined (Figure 25). While the distribution of the data points around the centroid was not random, clustering was only identifiable for the Aversive preference. The plots did not show clear differentiation between the clusters for Preferred and Intermediate preference levels.

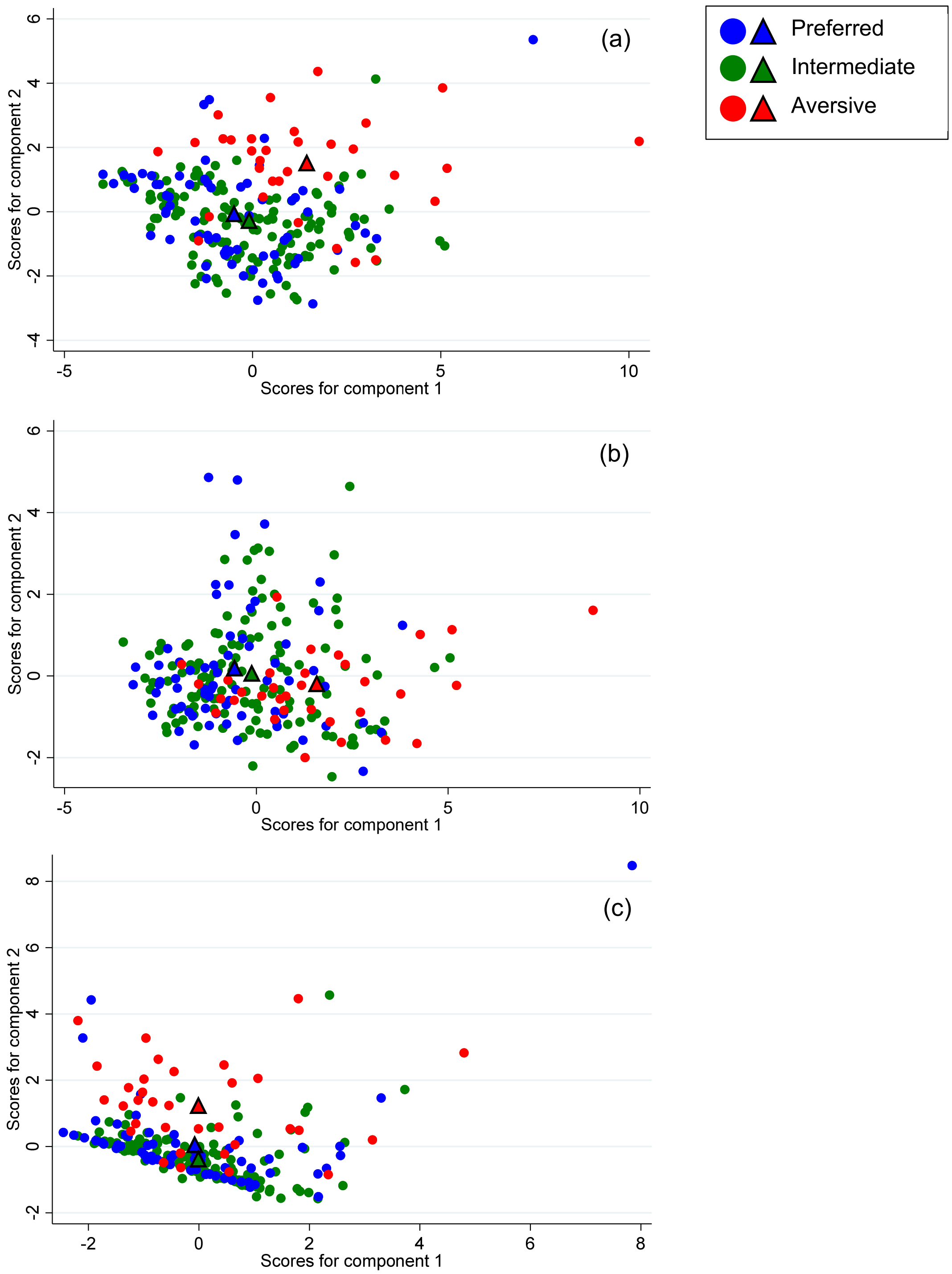


Figure 25 PCA score plots for (a) all 16 parameters that varied across the three preference levels during univariable analysis, (b) indicators of valence, and (c) indicators of arousal.

Dots represent individual scores; triangles represent the mean score within each preference level.

Examination of the PCA loading plots of the 16 parameters that varied across the three preference levels during univariable analysis (Figure 26) revealed differences in loading patterns between preference levels for nares flared, right ear side, and left or right ear back. Ears forward or right ear to the side have higher loadings and are not correlated with nares neutral for Preferred and Aversive but are correlated with nares flared for Preferred and Intermediate. Left or right ear back are related, and strongly negatively contribute to dimension 1 and dimension 2 for Intermediate preference level.

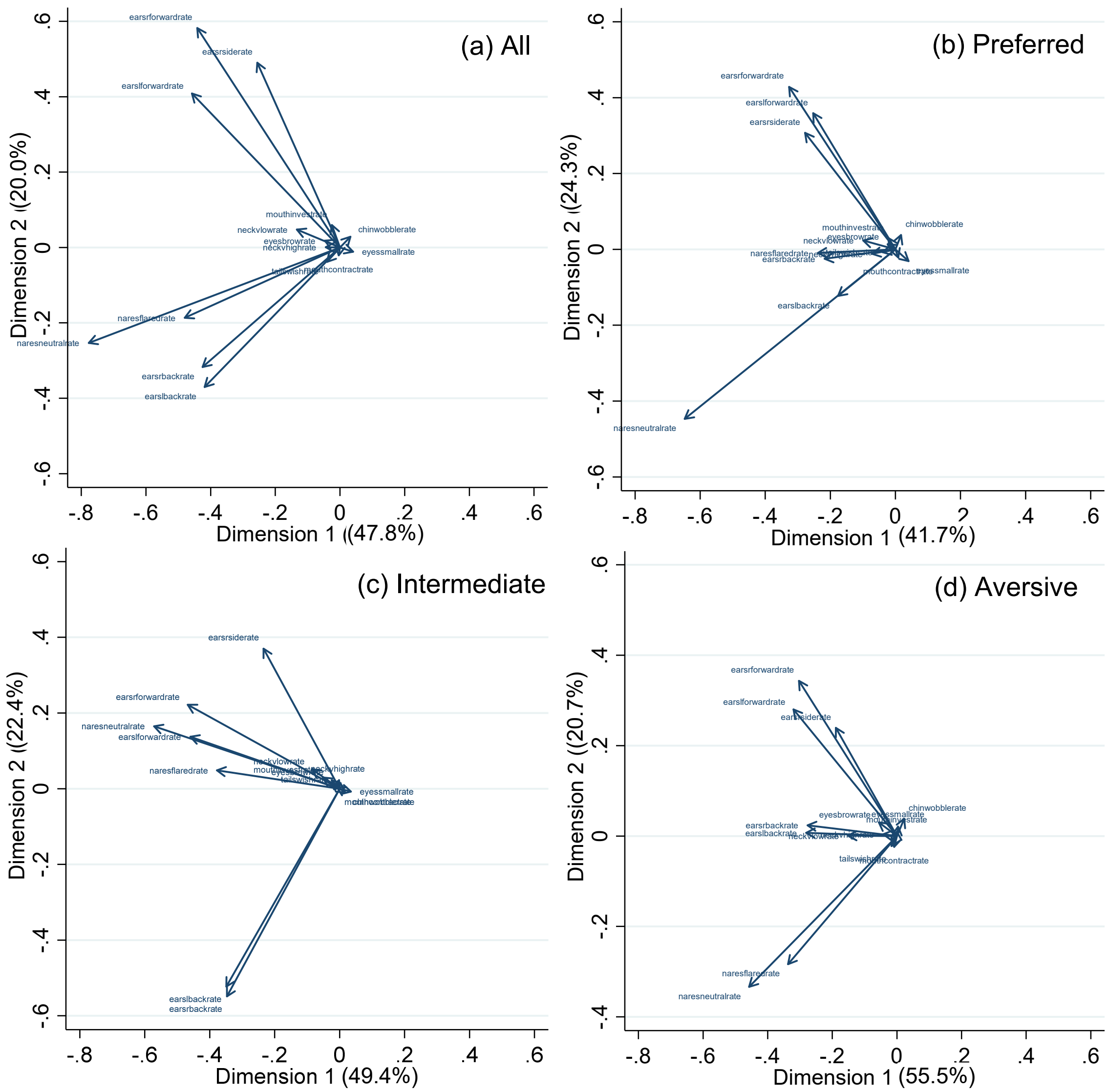


Figure 26 PCA loading plots for (a) all preference levels combined (explains 67.8% of total variance), (b) Preferred (explains 66.0% of total variance), (c) Intermediate (explains 71.8% of total variance), and (d) Aversive (explains 76.2% of total variance), preference levels.
Numbers in brackets indicate the proportion of total data variance explained by each dimension.

11.3.3 Linear regression analysis of PCA scores

The results of linear regression analysis for differences between preference levels in PCA scores of the 16 parameters that varied across all three preference levels during univariable analyses are presented in Table 34.

Principal components 1, 2, 3, and 5 differed between the Aversive and Preferred level of preference (PC1 1.93 (0.30) p <001, PC2 1.55 (0.20) p <0.001, PC3 -0.58 (0.22) p = 0.008, PC5 -0.51 (0.18) p = 0.005). There was no difference in PCA scores between the Preferred and Intermediate preference levels, although for PC1 and PC3 the p values tended towards the significance threshold.

Table 34 Results from a mixed linear regression model examining the effect of preference on the PCA scores for each observation.
Analysis is based on comparison to the Preferred category (referent). Where p ⩽ 0.05 results are shown in bold.

Component	Variable	Beta (SE)	p-value (Wald)
PC1 Nares-ears	Preferred	REF	
	Intermediate	0.40 (0.21)	0.062
	Aversive	**1.93 (0.30)**	**<0.001**
PC2 Ear forward/side	Preferred	REF	
	Intermediate	-0.21 (0.14)	0.147
	Aversive	**1.55 (0.20)**	**<0.001**
PC3 Oral-neck low-ear back	Preferred	REF	
	Intermediate	-0.27 (0.15)	0.075
	Aversive	**-0.58 (0.22)**	**0.008**
PC4 Neck high-oral-tail	Preferred	REF	
	Intermediate	-0.04 (0.15)	0.788
	Aversive	-0.14 (0.22)	0.524
PC5 Eye small-nares	Preferred	REF	
	Intermediate	-0.08 (0.13)	0.508
	Aversive	**-0.51 (0.18)**	**0.005**

11.4 Discussion

The current chapter examined clustering in the parameters that were found to vary across the three preference levels during univariable analysis in Chapters 7 to 9 using multivariate PCF analysis methods. Additionally, it explored how the individual parameters that were identified in Chapter 10 as indicators of the arousal and valence components of the underlying emotion, may be grouped. Relationships between parameters were compared for consistency with those expected from the univariable analysis. Using PCA and linear regression methods, it was determined whether the clusters differed by preference. A smaller set of parameters was identified that may serve as indices for the clusters, and that may be observed in future for simplified estimation of the relative emotional arousal and valence levels in horses.

Factor analysis using PCF clustered the behavioural data into five factors which explained two thirds of the variance. These were named Ears forward-side (Factor 1), Ears back-nares (Factor 2), Neck-tail (Factor 3), Blink-chin (Factor 4), and Eye-oral (Factor 5). The clustering of behaviours demonstrated by PCF are biologically plausible, when considered with regards to the behaviour demonstrated by the individual horses and their preferences. Additionally, comparison of the results from PCF1 with the results of the univariable analysis (Chapters 7 to 10) highlighted the nature of the relationships of the behaviours within the factors and consistency between the multivariate and univariable analyses for most parameters. For example, the rates of left and right ear back, nares flared, and nares neutral, which cluster together in Factor 2 and are related positively, were shown during univariable analyses to change similarly with preference (i.e., they all have an inverse relationship with valence level). Similarly, in Factor 4, within the factor blink and contracted lips have a positive relationship, and both have an inverse relationship with chin wobble, and this is consistent with findings of the univariable analyses (i.e., as valence increases from negative to positive, chin wobble increases, and blink decreases, and as arousal increases contracted lips increases). One parameter, neck very low, did not have a relationship within Factor 5 that was consistent with the findings of the univariable analyses. It is unknown why, but it is noted that this parameter's loading is near the $|r| > 0.40$ threshold for inclusion in Factor 5, and contributes more strongly to Factor 3, wherein the relationships are consistent with the univariable analyses. Perhaps its inclusion in Factor 5 is questionable. It is reassuring that the relationships expressed in the multivariate analysis largely align with those described in the univariable analyses.

It is interesting that some behaviours cluster together as expected, while others do not. For example, the neck behaviours are in the same factor (Factor 3), similarly for behaviours involving ears forward, ears back, and nares. However, the eye and oral behaviours occur two factors (Factors 4 and 5). This highlights the value of conducting multivariate analysis in addition to the univariable analyses to reveal unexpected patterns.

Despite the clustering of behavioural indicators, care should be taken when interpreting clusters using multivariate analysis. For Factor 1 (Ears forward-side), right and left ear forward clustered strongly. However, during univariable analysis right and left ear forward were found to be inverse indicators of arousal and valence, respectively. While these behaviours responded in the same direction, which fits with the PCF clustering, interpretation of the univariate analyses suggests that the individual behaviours may not indicate the same dimensions of emotion (i.e., arousal or valence). A previous study identified that individual ears may behave differently (Wathan et al. 2015), and can reflect differences in emotional laterality (Leliveld et al. 2013), with asymmetrical ear behaviour recorded in studies on emotion (Lansade et al. 2018). Thus, there is support in the literature for the findings from the univariate and PCF analysis for the differences in interpretation of left and right ear behaviour.

There are two studies on emotion in horses that have included multivariate analysis (Briefer et al. 2017; Lansade et al. 2018). In the first, PCA revealed that head high duration contributed moderately to two of the three components (collectively explaining 59% of the variance), and chewing contributed majorly to the third component in a study investigating horses' responses to audio recordings of positively or negatively valenced whinnies (Briefer et al. 2017). However, comparison with the current study is limited as insufficient detail is given on how the behaviours changed with the stimuli. These behaviours are similar to the neck high duration and licking and chewing rate behaviours of the current study, which were significant for the positive to negative (Preferred compared to Aversive) comparison. However, neither of these differed significantly across all three preference levels during univariable analyses and therefore were not included in the multivariate analysis of the current study. This highlights the care required during interpretation of the constituent parameters of components with the explanatory variable, in this case emotion, and the need for both univariable and multivariate analyses.

In the second study, PCA analysis based on Spearman's correlations of behaviours seen during two conditions in horses (standard and gentle grooming) conducted by Lansade et al. (2018) found that two factors explained 60% of the variance. The first component neatly, but perhaps artificially, differentiated the scores of the two treatment groups. They used a simpler, smaller dataset, with up to 20 freeze frames selected from the 10-minute video of each horse for a single session. The differences in study design (e.g., one replicate measured versus three, time sampling versus continuous/all occurrences behaviour measurement, two levels versus three, parallel versus crossover), smaller data set (27 observations versus 231), and lower number of parameters (12 versus 16) likely contribute to the reduced variance in the data compared to the current study. Additionally, there were differences in analysis methods that may also impact direct comparison between the studies. Despite these differences, some broad support for the current study can be found with the neck height, eye aperture, and mouth behaviours and their contribution to clusters.

Interestingly, despite finding no significant differences between treatment groups during univariable analysis, the authors found that ears pointing forward and ears pointing backwards contributed to the components (Lansade et al. 2018). White of the eyes contributed to the first component and was found to be statistically different between the treatment groups in univariable analyses, but the frequency was zero for both groups. This further reinforces the value of careful interpretation of multivariate analysis and utility of comparison with the univariable results.

The additional benefit of conducting exploratory factor (PCF) or component (PCA) analysis, is that single behaviours within each factor could be used as indices for that factor. In the current study, the subsets of behaviours that had been previously suggested to indicate valence and arousal were also analysed separately using PCF. Parameters that indicate arousal were found to reduce to two factors. These two factors explained two-thirds of the variance in the arousal specific PCF model (PCF2). The parameters with the highest loading value in each factor were right ear forward, and tail swish. Similarly, valence was found to reduce to four factors that explained approximately two-thirds of the variance in a valence specific PCF (PCF3). The parameters with the highest loading value in each factor were right ear back, small eye aperture, oral investigation, and blink. It may be that these six behaviours can be used as indices of each factor to estimate the relative arousal and valence levels more simply than measuring all sixteen parameters.

However, given that the tail swish behaviour only tended towards significance between Preferred and Intermediate levels, it may be prudent to replace tail swish with the behaviour with the next highest loading value in that factor, neck very high, as the index behaviour for Factor 2 in PCF3 when estimating arousal level. Similarly, as blink was removed from the PCA analysis due to not reaching threshold for inclusion, it may be prudent to replace blink with chin wobble as the index behaviour for Factor 4 in PCF2 when estimating valence level. Thus, the indices for estimating arousal would be neck very high and right ear forward, and the indices for estimating valence would be right ear back, small eye aperture, oral investigation, and chin wobble. Future research or practical application may be able to approximate the same results with reduction of the number of parameters measured to these few indices. Further research is required to confirm the findings.

Subsequent, regression analysis of the predicted PCA component scores revealed that clustering was evident for the Aversive level in four of the five components. However, differentiation between the Preferred and Intermediate levels only tended towards significance for two components, with no differences found in the remaining components. The findings from the linear regression analysis of the PCA scores, confirmed those from the visual examination of the PCA score plots. The results further support the findings from univariable analyses that the changes seen with highly contrasting emotional valence (Aversive versus Preferred) are more marked than those that have less contrast (Preferred versus Intermediate). The components or factors may be useful for differentiation of contrasting emotional experience, but not necessarily for identifying the more subtle differences between intermediate and more positive ones.

Limitations

Aside from the limitations of the data set (described in Chapter 10), one limitation of PCA and PCF analysis is that the data must be the same data type, and on the same scale (in this case continuous rate data) for all analysed parameters. Another limitation is that with PCA and PCF categorical data (e.g., horse, treatment, preference, replicate) cannot easily be included in the analysis with the continuous data. Both PCF and PCA analysis methods were used as each offered useful information within the constraints of the statistical software e.g., PCF rotation, PCA scores for regression analysis. The combination of techniques provides a more comprehensive view.

11.4.1 Conclusion

In conclusion, in situations of contrasting emotional experience it may be possible to estimate the relative valence level from observation of the rates of right ear back, small eye aperture, oral investigation, and chin wobble, and estimate the relative arousal level from observation of right ear forward and neck very high behaviours. Caution is advised when differentiating between more positive and intermediate emotional experiences, that is, where there is less contrast in valence and/or arousal.

The value of multivariate analysis for exploring relationships between multiple parameters and simplification of large datasets was demonstrated. Comparison with results from univariable analysis is useful to confirm expected relationships. Interpretation of the results of exploratory factor analysis with respect to the explanatory variables is aided by the findings from univariable analysis. This work forms a foundation for future research in the developing field of assessment of emotion.

Chapter 12 General discussion

This chapter opens with a discussion of the research approach and research issues. An alternative emotional framework is presented with integration of the findings of this research. A summary of the thesis and its contribution to the field is given. Practical applications of the findings are suggested, and future research directions proposed, before the concluding remarks.

12.1 Research approach

The research approach taken following the initial experiment (Chapter 4) involved two further experiments (Chapters 5-11). Preference testing of four stimuli was utilised to verify and order three levels of valence on an individual basis. Arousal was ordered in three levels by preference according to theorised salience. Non-invasive physiological and behavioural parameters were measured while horses were exposed to each of the four stimuli. Univariable analysis using mixed-effects models were used to identify parameters that varied with preference and the underlying acute emotional experiences. Interpretation of results identified indicators of the arousal and valence dimensions of emotional experience. Multivariate analysis using exploratory factor analysis was used to identify grouping of the parameters and to simplify the number of indicators.

Several assumptions underpin the research approach. First, a preferred stimulus is more desirable than the other stimuli and induces a more positive emotion (Humphrey 1972; Boissy et al. 2007b; Dawkins 2008). Second, the Spray stimulus is aversive and induces negative emotion (Visser et al. 2003; Lansade et al. 2010). Third, the other two stimuli are likely to induce emotion with a valence that falls between that of the most and least preferred stimuli (intermediate valence). Fourth, a horse's preferences were stable over time and location. Fifth, that there would be sufficient differences in the magnitude of acute emotion in order to detect significant changes in the measured parameters. Lastly, it was theorised that arousal could be ordered based on relative salience as shown by preference testing.

The first two assumptions were based on the literature thus there is confidence in these foundational premises of contrasting valence. Many differences were found on this basis. However, it must be considered that the preference categories represent relative levels of valence (i.e., more or less positive or negative), and that there is no universal positive, neutral, or negative point on an emotional valence scale. Thus, the findings of this research are relative within an individual observed experiencing contrasting emotional conditions. The intermediate valence assigned to the Intermediately preferred stimuli seems logical and fits with observation of horse's responses. It is recognised that the two stimuli may not have been regarded equally in terms of valence and/or arousal within the individual horses. This may have increased the variation in responses seen with the Intermediate preference level and made it harder to find significant differences between the Preferred/positive and Intermediate emotional experiences. This would have had a conservative effect on the results, making it possible that nonsignificant parameters may have been incorrectly rejected as indicators of emotion.

As expected, changes with positive emotional valence were more subtle and variable (Boissy et al. 2007b; Whittaker et al. 2019). Aside from the possibility that the nonsignificant parameters are not indicators of emotion, it is possible that that the chosen treatments may not have elicited sufficient contrast between the Preferred and Intermediate preference categories to reach the limit of detection of the measures and methods used. It may be that the sensitivity of the measures used for the nonsignificant parameters were too low, and/or a larger number of subjects and/or replicates may have been required, to detect more subtle changes. The effect of individual variation in response to the two intermediate treatments could also have overshadowed any real effects. This may be alleviated in future with use of a single intermediate treatment and/or horses that had the same preferences. However, limiting the number of stimuli may increase potential impacts of stimulus specificity. In future, use of more subjects and/or technologies, such as facial movement/expression recognition, or kinematics, and algorithms/artificial intelligence for automated data capture processing, could allow detection and differentiation on a finer level (Neethirajan et al. 2021).

Although preference stability was not tested over time and location during this thesis, and an effect cannot be ruled out, other research in horses has demonstrated persistence a year later (Lansade et al. 2018). Differences were found between the preference levels that were logical and support was found in the literature. However, if preferences had changed between the second and third experiments, the interpretation of the results by preference would be undermined. In that case, it is conceivable that the findings with the Aversive/negative condition may still be true, but attribution of Preferred/positive and Intermediate levels could be incorrect. Further research is suggested to corroborate the findings.

In the current study, heart rate did not differ across all preference categories, and thus could not be used to rank arousal level (Briefer et al. 2015b). Instead, the relative arousal of each preference category was determined theoretically, based on likely salience, survival impact, and previous literature. The Aversive stimulus would induce the highest arousal level, and the Intermediately preferred stimuli would induce the lowest. The Preferred stimulus may produce an arousal level somewhere between the Intermediate and Aversive stimuli. It is recognised that, while there is reasonable confidence in the method of differentiation of valence levels based on preference testing, arousal levels were ordered based on a theory which has not been empirically tested. If the theory is wrong, at the least, the indicators may still be used to differentiate emotional experience, but it may be uncertain what emotional dimension they represent.

No other studies have used three levels of valence and arousal, verified valence using an individual animal-based method, and ordered valence and arousal to the author's knowledge. Many studies have used two stimuli, and presumed the valence or arousal (Coulon et al. 2015; Lansade et al. 2018; Tamioso et al. 2018). Few studies have considered more than two stimuli, positive, negative and intermediate levels of valence and/or multiple levels of arousal in horses (Briefer et al. 2015b; Thorbergson et al. 2016; Janczarek et al. 2018b) or other species (Reefmann et al. 2009b, c; Reefmann et al. 2009a). Few studies have attempted to validate presumptive valence by scoring approach/avoidance behaviour (e.g. Lansade et al. 2018), or providing the animal with control/choice, e.g. voluntary grooming in sheep (e.g. Reefmann et al. 2009a). Few studies have attempted to categorise arousal, and this has been previously based on heart rate or heart rate variability (e.g. Briefer et al. 2015b; Lansade et al. 2018). Use of at least three levels of valence and/or arousal appears essential for attribution of responses to indicators of different levels of valence or arousal. The importance of independent verification of valence on an individual basis was highlighted when comparing the results of analysis by treatment and preference.

The current studies examined facial and body behaviours in greater detail than other studies, in stationary, unridden horses, using a cross-over design. Horses were loosely restrained to limit confounding effects of locomotion and enable detailed investigation of facial behaviours, with horses habituated to the experimental facility to reduce any effects of loose restraint. While the horses did not have choice about the application of the treatments, this is consistent with other studies in horses, and was consistent between the second and third experiments.

The behaviour sampling and recording methods differed from others in that rather than focal, scan, time sampling, arbitrary freeze-frame selection, or behaviour scoring, the current approach measured all occurrences and durations of behaviours from video recordings. This enabled detailed, accurate, less subjective measurements.

The analysis method differed to other studies. Some studies analysed data as average change from an individual or group average baseline. That method was not appropriate for experiment three as the baseline situation (No-stimulus) was also one of the treatments. There may also be issues with utilising a change from baseline method, as baseline may vary between and within horses, and the approach can misrepresent the impact of individual horse variability in both baseline and magnitude of response. For those reasons, the revised research approach used regression analysis models for data analysis. It is necessary to recognise that there is no universal neutral valence and/or arousal situation that can be applied equally across horses for between-subject comparison. There is also no gold standard indicator of emotional valence or arousal. Therefore, results are relative to the other preferences or treatments, rather than absolute.

Using this research approach, body behaviour, facial movements, and physiological parameters were found to vary across treatment conditions. Analysis by preference and treatment highlighted the need to consider individual regard (preference) for the stimuli. What horses like may be unique to the individual and cannot be assumed. Thus, emotion inducing stimuli need to be tailored to the individual, and independent objective verification of the likely emotional valence is required. Alternatives to the preference testing method used in this thesis might be pairwise preference testing, or other measures of approach and avoidance behaviour. Whichever method is used to verify the emotional valence, the underlying variables cannot then be said to be indicators of emotion, otherwise a circular argument is formed.

The species used in the model, horses, are utilised as a companion as well as a production animal domestic species. At the time the field work was undertaken (2010-2011), positive emotion had not been investigated in horses (to the author's knowledge). It is suggested that the research approach and study designs were suitable for investigation of indicators of emotion in horses and could be a model applied to other domestic species, with some species-specific modifications.

12.2 Research issues

Along with the methodological considerations and limitations discussed in preceding chapters, some important issues were highlighted by this research work. These include aspects of study design, such as the nature of the emotion inducing stimuli, verifying presumptions of emotional valence by 'asking the animal', and comparing multiple levels of valence and/or arousal; clarity issues, such as explicitly stating foundational assumptions, defining terminology, and ethograms; as well as consideration of the sensitivity and specificity of potential indicators of emotion and their confounding factors. In future, technology may help to address sensitivity of measures (e.g., facial movement assessment software), and reduce potential confounds from human presence and effects of sampling/data collection (e.g., remote biosensors, automation).

Difficult questions remain, such as the extent to which cognition and consciousness influence emotion in animals, the impact of individual traits, and the types and timeframes of emotion in animals (e.g., what constitutes acute/momentary emotion, emotional state or mood, or life-long emotion). The issues of potential interaction between emotional valence and arousal, and the relativity of both dimensions of emotion due to the lack of a universal neutral or zero point, are as yet unsolved. There are ethical and safety concerns around induction of negative emotion in the research context, and long-term consequences on future behaviour. There is also potential for confounding effects of negative emotion, such as overshadowing the induction of positive emotion during research, in the short term. Although no indicators have proven incontrovertible enough to enable a gold standard for future comparison, cross-study comparison and progression of the research field could be aided by a degree of standardisation across study design, methods of emotion induction, parameters, terminology, and analysis methods.

12.3 An alternative emotional arousal-valence framework

The arousal versus valence emotional framework is a commonly used foundation for empirical research on subjective emotion (Russell 1980; Knutson et al. 2002; Paul et al. 2005; Mendl et al. 2010). It allows for the conceptualisation of the arousal and valence dimensions of emotion along continua, from low arousal to high arousal on the vertical axis, and negative to positive emotional valence on the horizontal axis (Figure 1, Figure 2).

Controlling for arousal may be key to unravelling the arousal-valence interplay and elucidating specific identifiers of emotional valence. Just as there is no gold standard for determination of emotional valence in animals, there is no agreed gold standard for assessment of emotional arousal. Many physiological parameters have been suggested to represent arousal, but these have often been investigated in the context of (likely) negative emotional valence (e.g., pain, fear, frustration), and can be influenced by other factors (e.g., locomotion/movement, context/stimulus). The difficulty lies in finding a measure of arousal that is sensitive across the continuum of negative-positive valence, is specific to emotional arousal, and reliable across stimuli and contexts.

Previously, attempts have been made to rank emotional experience based on heart rate as a measure of arousal (Briefer et al. 2015b; Briefer et al. 2017). Although there were problems with the methodology and application that limited the utility of the results, the concept and analytical methods, if combined with a means of more objectively assigning valence, could be a useful approach in future research where there is sufficient variation in heart rate across valence levels. However, in the current study, that approach was not useful, as although heart rate varied between the most positive and negative emotion conditions, it did not vary between the positive and intermediate conditions. Instead, arousal was ordered based on the theorised relative salience of a preferred, intermediate, and aversive stimulus. Validation of this theory using empirical research would be useful in future research, although challenging to design without a gold standard comparator.

The graphical representations of the existing two-dimensional arousal versus valence framework suggest that a neutral or zero valence point exists which coincides with a mid-point between low and high arousal (Knutson et al. 2002; Mendl et al. 2010). According to the graphics, it would seem possible to have a low or high arousal negative, a low or high arousal positive emotion, but also a low or high arousal neutral emotion (Figure 1,Figure 2). It is difficult to envisage a neutrally valenced emotional experience that is highly arousing. A neutral valence implies that a particular stimulus is inconsequential. However, if arousal is increased, then the stimulus is meaningful.

In this thesis, it was posited that, due to salience and survival impact, an aversive stimulus was likely to stimulate a higher level of arousal than that due to an intermediately preferred one, with a more preferred stimulus having an arousal level in between. In extension of that reasoning, it is also suggested that, generally, an emotion cannot be arousing and have a neutral valence.

Alternatively, over the continuum of valence, emotions could be thought to be located on a U-shaped curve, with a trough at the theoretical neutral valence point. This idea is presented graphically in Figure 27 and is termed the 'U shaped emotional arousal-valence dimension framework'. As the degree of positive or negative valence increases, so does the level of arousal. Examples of generally putatively positive and negative emotional experiences, for non-impoverished animals, are shown in blue text. Approximate, relative locations of the three preference categories used in this thesis are overlayed in red text. The concept may prompt further discussion around the frameworks for understanding emotions, and perhaps be a useful construct in future research on emotion.

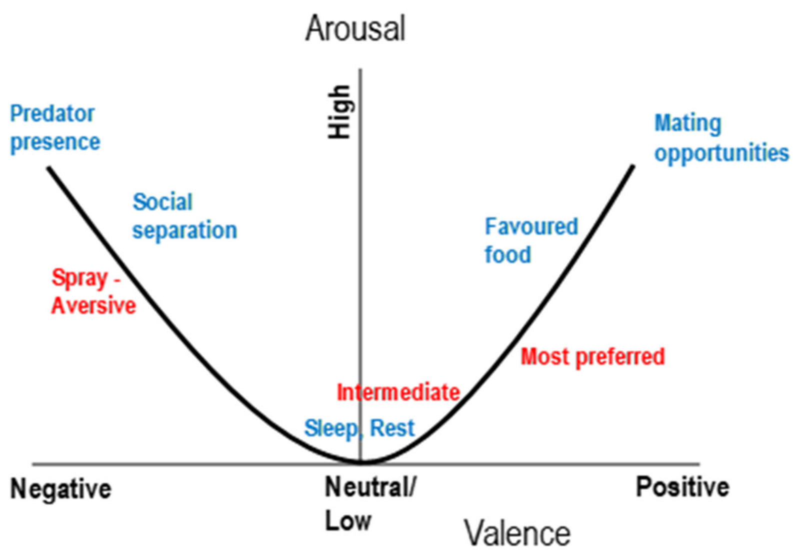


Figure 27 Graphical representation of a modified arousal versus valence framework, termed the 'U shaped emotional arousal-valence dimension framework'.
Emotions are on a U-shaped curve where low arousal coincides with neutral valence. Examples of emotional experiences that may be located on different points of the curve are given in blue text.
The three preference categories used in this research are overlayed in red text, showing approximate relative position.

This new framework conceptualises emotion in a two-dimensional space. It may be in future that other dimensions are added, such as time/duration, to conceptualise the magnitude of emotional response, and/or depict the categorisation of momentary/acute emotion, emotional state/mood, and lifelong emotion/affective happiness. Further, as suggested by (Mauss et al. 2009), the addition of an approach-avoidance dimension may help with reconciling the discrete and dimensional perspectives of emotion, and 'mapping' of discrete emotions.

The indicators of emotional arousal and valence as determined by the research presented in this thesis are overlayed onto the graphic in Figure 28. Increases in the rates of small eye aperture, chin wobble, or oral investigation indicate a more positive valence, i.e., shift along the curve to the right. Increases in the rates of neck very low, left ear forward, left ear back, right ear back, blink, angled eyebrow, nares flared, or nares neutral indicate a more negative valence, i.e., shift along the curve to the left. Increases in the rates of neck very high, tail swish, or probability of contracted lips indicate an increase in arousal, i.e., a shift up the curve. Increases in the rates of right ear forward or right ear side indicate a decrease in arousal, i.e., a shift down the curve. Decreases in the rates/occurrence of the indicator behaviours have the opposite meaning.

The underlined behaviours are the index behaviours for the clusters determined during multivariate analysis (Chapter 11). It may be possible to estimate the relative arousal and valence from this subset of indicators.

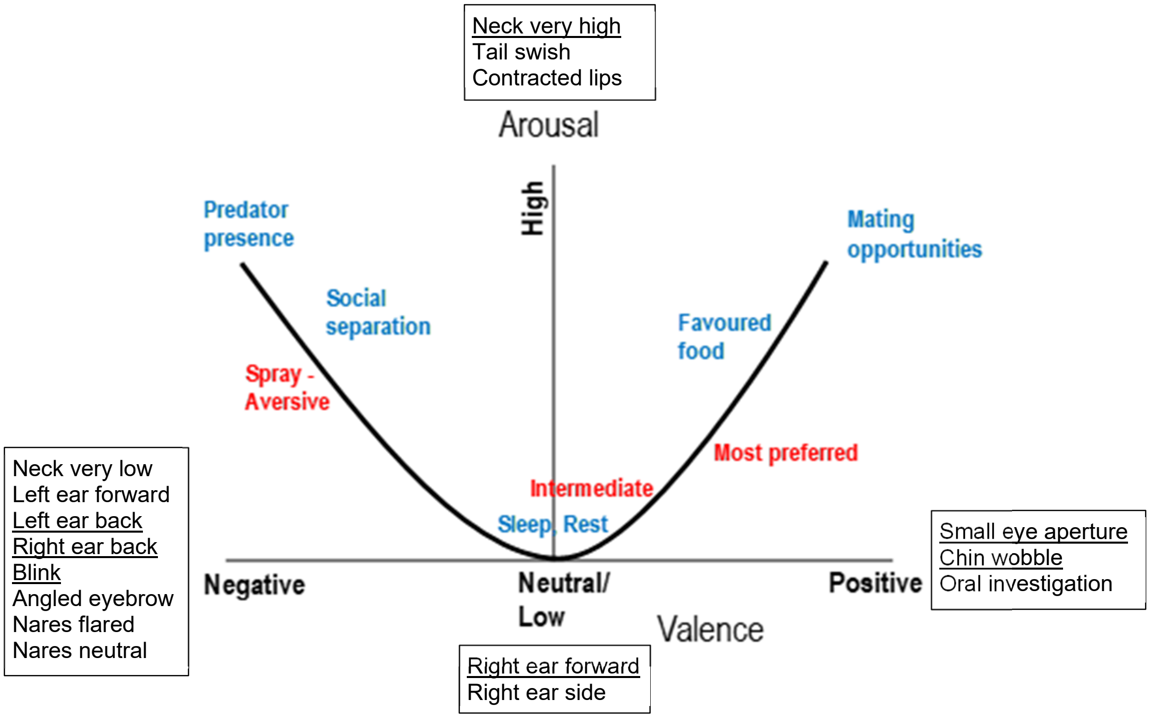


Figure 28 U shaped emotional arousal-valence dimension framework, with indicators of arousal and valence overlayed.
Increases in the rates of the behaviours (or in the case of contracted lips, the probability of occurrence) indicate change in arousal or valence in the directions as shown. Conversely, decreases in the behaviours indicate a change in the opposite direction. The underlined behaviours are the index behaviours for the clusters determined during multivariate analysis. It may be possible to estimate the relative arousal and valence from this subset of indicators.

12.4 Thesis summary and contribution

The object of this thesis was to advance the knowledge of assessment of acute emotion in animals, using the horse as a model. The aim was to investigate if emotional experiences of horses were reflected in behavioural and physiological parameters to determine their utility as indicators of emotion, and in particular, positive emotion.

The evidence for the existence of emotion, including positive emotion, in animals, was examined, providing a foundational premise for the work. The empirical research on methods that may be useful for inducing emotion, and prospective measures for identification of positive emotion, in horses and other species, was reviewed. From this base, stimuli likely to induce emotion of differing valence and parameters with possible use as indicators of emotion were selected. A series of three experiments were conducted to investigate potential non-invasive, animal-based indicators of positive emotion in horses. The first experiment revealed issues with the research approach, which were addressed via experiments two and three using novel methods.

The first review (Chapter 2) concluded that emotions, positive and negative, are likely to be experienced by animals. Research on assessment of acute positive emotion in animals is required (Boissy et al. 2007a; Fraser 2008; Yeates et al. 2008; Yeates 2011). In Chapter 3, review of the literature identified that limited research has been conducted expressly on positive emotion in horses. The progress of empirical research on detecting positive emotion in animals is hampered by differing views on the function of emotion, frameworks for studying emotion, and specific types of emotions theorised to exist in animals. Parameters that could be considered for further investigation as potential indicators of emotion in animals were identified.

The first experiment, presented in Chapter 4, involved the collection of physiological and gross behavioural data from horses undergoing presumed positive and negative treatments. A large degree of variability was found, and what individual horses liked and disliked was unique and could not be assumed. From this realisation, a new research approach was formed.

In the second experiment, presented in Chapter 5, a method of preference testing, termed the 'location-associated simultaneous preference test', was designed to allow horses to indicate their preferred stimulus, by choosing to spend more time in an area associated with that stimulus. A preference for the offered stimuli was displayed by half of the horses. Horses' preferences differed, with the majority preferring having a person nearby, to wither grooming, or to being left alone. The horses that demonstrated a reliable choice were then used in the next experiment. The preference test results provided a novel way for the emotional valence induced by the stimuli to be ordered on an individual basis.

The last experiment involved exposing those horses to each of the four stimuli while recording numerus non-invasive physiological and behavioural parameters. The general introduction and methodology for the experiment was presented in Chapter 6. Comparison of selected physiological (Chapter 7), body behavioural (Chapter 8), and facial movement (Chapter 9) responses of horses experiencing a most preferred, intermediately preferred, and an aversive treatment, were made in order to identify indicators of underlying acute emotional experience.

The findings from Chapters 7-9 were assimilated in Chapter 10, a novel theory on salience and arousal was applied, and the results interpreted with respect to the arousal and valence dimensions of emotion. These were discussed in Chapter 10, along with the limitations of the experiment. To further understand the findings, multivariate analysis of the parameters that varied across the three preference levels during univariable analysis was performed (Chapter 11), attempt was made to identify clustering and to refine the number of parameters for future work.

In total, 45 parameters differed by preference. Information about horses' emotional experience was reflected in changes to heart rate, heart rate variability, eye temperature, and behaviours involving legs, neck, tail, ears, eyes, mouth, chin, and nares. As expected, it was easier to differentiate large contrasts of emotion such as that between positive and negative (44 parameters), but harder to differentiate positive and intermediate emotional experience (17 parameters, smaller magnitude of difference). Sixteen parameters differed across all three preference categories. They were all behavioural, with the majority being facial movements.

In this study, emotional valence in horses may be indicated by the frequency of neck very low, left ear forward, left and right ear back, blinking, angled eyebrow, nares flared, nares neutral, small eye aperture, chin wobble, and oral investigation behaviours. Emotional arousal in horses may be indicated by the frequency of behaviours such as neck movements to the very high position, right ear forward and to the side, tail swishing, and the probability of contracted lips behaviours. The findings suggest that horses may convey information about emotional experience through facial expression.

These studies explored the assessment of emotion, including positive emotion, an emerging area of animal welfare science, in detail in a novel way in an understudied species (Hall et al. 2018). A substantial amount of information was gathered from review and investigation of numerous parameters. At the time it was conducted, this work was, to the author's knowledge, the first investigation of facial expression, and eye temperature, as potential indicators of positive emotion in horses. The findings provide valuable insight, contributing substantially to the body of knowledge, and forming a foundation for further research. The thesis identified observable, non-invasive behavioural indicators for determination of emotional arousal and valence that may be suitable for use in clinical, research, and field settings.

Many research issues were identified, and resolution attempted. The use of three levels of independently verified valence, with a mechanism for interpretation of responses with respect to the arousal and valence dimensions of emotion, in horses, and such in-depth investigation of body and facial behaviours, to the authors knowledge, has not been conducted before. The thesis contributes a research approach that may be applicable for use in future research on emotion in horses and other species.

Lastly, the posited alternative framework for conceptualising emotion, the ‘U shaped emotional arousal-valence dimension framework’, presented in Chapter 12.3, challenges and extends current thinking in the field of emotion. Whether the idea stands up in the face of further research and discussion remains to be seen, but at the least, it may be a stimulus for future advancement of the field.

12.5 Practical applications of indicators of emotion

Equestrians may not correctly interpret behaviour and the underlying emotion in horses (Lansade et al. 2019; Bornmann et al. 2021). The indicators identified in this body of work may be used in the assessment and improvement of the welfare of horses (Stratton et al. 2014). Situations that induce signs of negative emotion could be minimised, and situations that elicit signs of positive or intermediate emotion could be encouraged during management and training, thus improving the welfare of horses in the field setting (Hall et al. 2018; Kelly et al. 2021). Understanding emotion is also crucial to effective treatment of problems in veterinary behavioural medicine (Lezama-García et al. 2019).

Injuries sustained when dealing with horses are a significant cause of serious morbidity (Guyton et al. 2013; Jones et al. 2018). Recognition of horse behaviours indicative of increased arousal levels, particularly with negatively valenced experiences, may be useful in safeguarding human and horse safety (Hall et al. 2018). Education on horse behaviour may improve safety (Northey 2003). This work can be used to support evidence-based education on horse behaviour, and improved safety during interactions with horses through observation of behaviours and facial expressions.

It must also be remembered that, whilst particular behaviours and facial expressions were associated with differences in emotional arousal and valence in this study, they are not pathognomonic to the ascribed emotion. Human facial movements may express emotion, but are not specific for it; they often communicate other information and can be context dependent (Barrett et al. 2019). The same may apply in animals, which highlights the need for careful field interpretation with other signs, context, antecedent events, and consequences.

Horses' preferences differ. The identification of a horse's likely preferred stimulus, either from preference testing or from observation of behaviours found in this work to be associated with positive emotion, may enable environmental enrichment activities to be based on individualised, objective assessment, rather than general assumptions (Campbell et al. 2021). Counterconditioning, involving concurrent exposure to a positive stimulus, may decrease the negative emotional response when performing necessary, but aversive, tasks with horses. The effectiveness of counterconditioning efforts may also be enhanced by individualising the positive stimulus. For example, some horses find wither grooming more or less preferable than others.

12.6 Future directions

Further research is suggested to corroborate and expand on the findings. Future investigations of indicators of emotion could use subjects with the same treatment preferences and use a single intermediately preferred treatment. Ideally, valence or arousal would be controlled and tested separately, although this may be difficult to achieve in reality. Theoretically, stimuli used to induce negative, intermediate, and positive emotion would have the same level of arousal and use the same type of stimulus. However, the inclusion of a range of stimulus types reduces the impact of stimulus specificity on outcomes. Use of a larger number of subjects, and treatments that have a greater disparity between preferred and intermediate categories, might also be beneficial in future research.

Examination of the duration of emotion, and other contexts with different stimuli and levels of valence and arousal would be valuable. Repetition in a field setting, or other context, that minimises potential impacts from humans, perhaps using voluntary engagement with a positive emotion inducing stimulus (for example automatic rotating brushes), may yield interesting results.

It may be of interest to examine naturally occurring behaviours to determine the validity of the results in the field setting, and/or to elucidate communicative functions of the facial expressions between conspecifics, or possibly between species. For example, is the display of particular facial behaviours associated with continuation or cessation of mutual grooming.

Further examination of the convergence of different measures of emotion in animals is suggested in future. A within-individual, rather than between-individual, study design is suggested to most appropriate. Low correlations among multiple measures have been found in humans and horses, which may point to the existence of moderator variables that impact convergence across measures of emotion (Mauss et al. 2009; Kelly et al. 2021).

Research investigating the attenuation of a negative emotional response, by co-administration of a positive emotion inducing stimulus (counterconditioning), could have practical implications for horse welfare and human safety. It may be that the magnitude of reaction to a negative emotional experience is reduced, or that recovery post-exposure is enhanced.

Extension of this research could involve assessment of the use of behavioural observation by animal caretakers in the field, for welfare and safety applications. It would be useful to first determine whether observation by horse caretakers, of the indicators of emotion, as described in this thesis, leads to correct identification of a horse's preference for a treatment and/or emotional valence and arousal in the applied setting. It may be that some indicators are easier and more accurate for use in the applied setting by caretakers than others.

There may be benefits to broadening the mainly quantitative research approaches used in investigation of emotions to include qualitative methods, such as descriptive observations, proxy-report, and qualitative behavioural assessment, or other narrative data (Kremer et al. 2020). It may be that characteristic patterns are apparent when several research methods are considered together, though this level of analysis may be challenging.

Research suggests that some equestrian's interpretation of global horse behaviour and perceptions of emotional experience may be inaccurate (Bornmann et al. 2021). However, qualitative behavioural assessment (QBA) is a method that could be used for assessing the correlation of gross behaviour interpretation by horse caretakers with the results from the current studies. In the not-too-distant future, through the use of machine learning (artificial intelligence) and automated facial expression measurement, it is possible to imagine a phone app for use by animal caretakers that interprets facial expression in terms of emotion (Andersen et al. 2021).

Although invasive, future investigation of potential neurohormonal correlates of emotion may help to provide further support for interpretation of the minimally invasive indicators of emotion. Secretion of neuropeptides such as endorphins, oxytocin, and prolactin are known to be associated with modulating emotional states in humans, and are maximal in blood draining the pituitary gland. The unique anatomy of horses allows for non-surgical collection of pituitary venous blood, which may prove to give more reliable measurements of neuropeptides than peripheral blood samples. Functional brain imaging, although challenging in large species such as horses, may also provide valuable correlations with minimally invasive measures. These methods, along with cross-species investigations, may help connect the homological dots for emotion and indicators of emotion.

12.7 Final thoughts

We observe the behaviour of animals and utilise our interpretation of it during everyday interactions. Although seemingly intuitive and simple, the process of providing objective metrics to this capacity for human observation of animal emotional experience is surprisingly complex.

"*Sentience is not just a 'hard problem'. It is the hardest problem in the whole of biology*" - Dawkins (2006)

Good welfare for animals includes the experience of positive emotions, as well as minimisation of negative emotions. Objective, minimally invasive indicators of positive emotion are desired to support assessment of animal welfare and welfare improvement efforts. Yet, identifying subjective emotion in animals is a challenging issue, and knowledge regarding positive emotion is vastly incomplete for many species.

Darwin suggested in 1872, in his seminal book The Expression of the Emotions in Man and Animals, that animal's facial expressions and postures can signal their emotional experience (Appleby et al. 2018). This body of work demonstrated that horses' behaviour and facial movements differ when presented with situations of contrasting emotional valence and arousal. This new work provides evidence that horses display facial expressions and confirms that facial behaviour in animals is valuable in assessment of positive and negative emotions.

Gaps in knowledge were identified and issues with emotion research were uncovered. A new approach to the problem of animal-based verification of valence, the 'location-associated simultaneous preference test', was developed. A new perspective to teasing apart the arousal and valence dimensions of emotion was described, and an alternative 'U shaped emotional arousal-valence dimension framework' put forward. It is possible that this approach may be applied in future research on emotion in horses and other species.

Investigation of many behavioural and physiological parameters revealed 45 that differed with contrasting emotional experience in horses. Sixteen behaviours were identified as indicators of emotional valence or arousal. These behaviours were clustered into five factors. The rates of right ear back, small eye aperture, oral investigation, and chin wobble may be used as indices of these factors to estimate the relative valence level. The rates of right ear forward and neck very high behaviours may estimate the relative arousal level.

However, all indicators of subjective emotion are indirect, and none are pathognomonic for emotion. It seems likely that no single measure will be suitable for assessing emotion in animals, and the use of a single measure may lead to erroneous conclusions. Consideration of multiple indicators, the context, confounding factors, and stimulus specific effects, is required during cautious interpretation.

To conclude, these studies may inform the assessment and promotion of good animal welfare and debate surrounding assessment of what activities horses do and do not enjoy. There may be a balance of positive to negative emotions which make negative experiences more acceptable for horses and/or to humans. Where and how the balance of acceptability is determined may be controversial. The arousal, valence, magnitude, and duration of emotional responses are all factors to be contemplated.

"*The more I learn, the more I realise how much I don't know.*" – Albert Einstein

References

Ahloy-Dallaire J, Espinosa J, Mason G. 2018. Play and optimal welfare: Does play indicate the presence of positive affective states? Behavioural Processes. 156:3-15.

Alberghina D, Rizzo M, Piccione G, Giannetto C, Panzera M. 2017. An exploratory study about the association between serum serotonin concentrations and canine-human social interactions in shelter dogs (Canis familiaris). Journal of Veterinary Behavior. 18:96-101.

Allen E, Sheaffer C, Martinson K. 2013. Forage Nutritive Value and Preference of Cool-Season Grasses under Horse Grazing. Agronomy Journal. 105(3):679-684.

Alvarez-Bolado G, Celio MR. 2016. The ventrolateral hypothalamic area and the parvafox nucleus: Role in the expression of (positive) emotions? Journal of Comparative Neurology. 524(8):1616-1623.

Andersen PH, Broomé S, Rashid M, Lundblad J, Ask K, Li Z, Hernlund E, Rhodin M, Kjellström H. 2021. Towards Machine Recognition of Facial Expressions of Pain in Horses. Animals. 11(6):1643.

Appleby MC, Olsson IAS, Galindo F, editor(s). 2018. Animal Welfare. Wallingford, Oxfordshire; Boston, MA: CABI.

Appleby MC, Mench JA, Olsson IAS, Hughes BO, editor(s). 2011. Animal welfare. 2nd ed. Wallingford, Oxfordshire ; Cambridge, (MA): CABI.

Arnold NA. 2005. A study of the behavioural and physiological responses of the dairy cow (Bos taurus) to selected environmental features of the milking facility. Melbourne Australia: The University of Melbourne

Arnold NA, Ng KT, Jongman EC, Hemsworth PH. 2008. Avoidance of tape-recorded milking facility noise by dairy heifers in a Y maze choice task. Applied Animal Behaviour Science. 109(2-4):201-210.

Ashley FH, Waterman-Pearson AE, Whay HR. 2005. Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. Equine Veterinary Journal. 37:565-575.

Austin NP, Rogers LJ. 2012. Limb preferences and lateralization of aggression, reactivity and vigilance in feral horses, Equus caballus. Animal Behaviour. 83(1):239-247.

Bachmann I, Bernasconi P, Herrmann R, Weishaupt MA, Stauffacher M. 2003. Behavioural and physiological responses to an acute stressor in crib-biting and control. Applied Animal Behaviour Science. 82(4):297-311.

Balcombe J. 2009. Animal pleasure and its moral significance. Applied Animal Behaviour Science. 118(3-4):208-216.

Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. 2019. Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest. 20(1):1-68.

Bartolomé E, Sánchez MJ, Molina A, Schaefer AL, Cervantes I, Valera M. 2013. Using eye temperature and heart rate for stress assessment in young horses competing in jumping competitions and its possible influence on sport performance. Animal. 7(12):2044-2053.

Basile M, Boivin S, Boutin A, Blois-Heulin C, Hausberger M, Lemasson A. 2009. Socially dependent auditory laterality in domestic horses (Equus caballus). Animal Cognition. 12(4):611-619.

Beausoleil NJ. 2006. Behavioural and physiological responses of domestic sheep (Ovis aries) to the presence of humans and dogs [Doctoral]. Massey University

Bermond B. 2001. A neuropsychological and evolutionary approach to animal consciousness and animal suffering. Animal Welfare. 10:S47-S62.

Berns GS, Brooks AM, Spivak M. 2012. Functional MRI in Awake Unrestrained Dogs. PLOS ONE. 7(5):e38027.

Beshkar M. 2008. Animal consciousness. Journal of Consciousness Studies. 15(3):5-33.

Blokhuis HJ, Veissier I, Miele M, Jones B. 2010. The Welfare Quality® project and beyond: Safeguarding farm animal well-being. Acta Agriculturae Scandinavica, Section A - Animal Science. 60(3):129-140.

Boissy A, Arnould C, Chaillou E, Desire L, Duvaux-Ponter C, Greiveldinger L, Leterrier C, Richard S, Roussel S, Saint-Dizier H et al. 2007a. Emotions and cognition: a new approach to animal welfare. Animal Welfare. 16:37-43.

Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, Winckler C, Forkman B, Dimitrov I, Langbein J et al. 2007b. Assessment of positive emotions in animals to improve their welfare. Physiology & Behavior. 92(3):375-397.

Bornmann T, Randle H, Williams J. 2021. Investigating Equestrians' Perceptions of Horse Happiness: An Exploratory Study. Journal of Equine Veterinary Science. 104:10.

Bowman A, Scottish S, Dowell FJ, Evans NP. 2015. ‘Four Seasons’ in an animal rescue centre; classical music reduces environmental stress in kennelled dogs. Physiology & Behavior. 143(0):70-82.

Brando S, Buchanan-Smith HM. 2018. The 24/7 approach to promoting optimal welfare for captive wild animals. Behavioural Processes. 156:83-95.

Bremhorst A, Sutter NA, Würbel H, Mills DS, Riemer S. 2019. Differences in facial expressions during positive anticipation and frustration in dogs awaiting a reward. Scientific Reports. 9(1):19312.

Briefer EF, Tettamanti F, McElligott AG. 2015a. Emotions in goats: mapping physiological, behavioural and vocal profiles. Animal Behaviour. 99:131-143.

Briefer EF, Vizier E, Gygax L, Hillmann E. 2019. Expression of emotional valence in pig closed-mouth grunts: Involvement of both source- and filter-related parameters. Journal of the Acoustical Society of America. 145(5):2895-2908.

Briefer EF, Maigrot AL, Mandel R, Freymond SB, Bachmann I, Hillmann E. 2015b. Segregation of information about emotional arousal and valence in horse whinnies. Sci Rep. 4:9989.

Briefer EF, Mandel R, Maigrot A-L, Briefer Freymond S, Bachmann I, Hillmann E. 2017. Perception of emotional valence in horse whinnies. Frontiers in zoology. 14:8-8.

Briefer Freymond S, Briefer EF, Zollinger A, Gindrat-von Allmen Y, Wyss C, Bachmann I. 2014. Behaviour of horses in a judgment bias test associated with positive or negative reinforcement. Applied Animal Behaviour Science. 158:34-45.

Broom DM. 2010. Cognitive ability and awareness in domestic animals and decisions about obligations to animals. Applied Animal Behaviour Science. 126(1-2):1-11.

Brosschot JF, Thayer JF. 2003. Heart rate response is longer after negative emotions than after positive emotions. International Journal of Psychophysiology. 50(3):181-187.

Bruce LL, Neary TJ. 1995. The limbic system of tetrapods - a comparative-analysis of cortical and amygdalar populations. Brain Behavior and Evolution. 46(4-5):224-234.

Burani C, Barnard S, Wells D, Pelosi A, Valsecchi P. 2020. Using judgment bias test in pet and shelter dogs (Canis familiaris): Methodological and statistical caveats. PLOS ONE. 15(10):e0241344.

Burgdorf J, Panksepp J. 2006. The neurobiology of positive emotions. Neuroscience and Biobehavioral Reviews. 30(2):173-187.

Burger D, Meuwly C, Thomas S, Sieme H, Oberthuer M, Wedekind C, Meinecke-Tillmann S. 2018. Cycle-specific female preferences for visual and non-visual cues in the horse (Equus caballus). Plos One. 13(2).

Çakiroğlu D, Meral Y, Sancak AA, Çifti G. 2007. Relationship between the serum concentrations of serotonin and lipids and aggression in dogs. Veterinary Record. 161(2):59-61.

Calvi E, Quassolo U, Massaia M, Scandurra A, D'Aniello B, D'Amelio P. 2020. The scent of emotions: A systematic review of human intra- and interspecific chemical communication of emotions. Brain and Behavior. 10(5x):e01585.

Campbell DLM, Lee C. 2021. A Perspective on Strategic Enrichment for Brain Development: Is This the Key to Animal Happiness? Frontiers in Veterinary Science. 8.

Canali E, Keeling L. 2009. Welfare Quality®project: from scientific research to on farm assessment of animal welfare. Italian Journal of Animal Science. 8(sup2):900-903.

Cannas S, Palestrini C, Canali E, Cozzi B, Ferri N, Heinzl E, Minero M, Chincarini M, Vignola G, Dalla Costa E. 2018. Thermography as a Non-Invasive Measure of Stress and Fear of Humans in Sheep. Animals. 8(9):146.

Carreras R, Arroyo L, Mainau E, Valent D, Bassols A, Dalmau A, Faucitano L, Manteca X, Velarde A. 2017. Can the way pigs are handled alter behavioural and physiological measures of affective state? Behavioural Processes. 142:91-98.

Catalano DN, Sheaffer CC, Grev AM, DeBoer ML, Martinson KL. 2019. Yield, Forage Nutritive Value, and Preference of Legumes under Horse Grazing. Agronomy Journal. 111(3):1312-1322.

Christensen JW, Keeling LJ, Nielsen BL. 2005. Responses of horses to novel visual, olfactory and auditory stimuli. Applied Animal Behaviour Science. 93(1-2):53-65.

Christensen JW, Zharkikh T, Ladewig J. 2008. Do horses generalise between objects during habituation? Applied Animal Behaviour Science. 114(3-4):509-520.

Church J, Cook N, Schaefer A. 2009. Recent Applications of Infrared Thermography for Animal Welfare and Veterinary Research: Everything from Chicks to Elephants.

Clark JA. 2010. Relations of homology between higher cognitive emotions and basic emotions. Biology & Philosophy. 25(1):75-94.

Cook N. 2005. Stress and pain assessment of velvet antler removal from Elk (Cervus elaphus canadensis) and Reindeer (Rangifer tarandus). Online Journal of Veterinary Research. 9(1):13-25.

Cook N, Schaefer AL, Church J. 2006. Nutritional therapy modulates stress responses of elk (Cervus elaphus canadensis) to removal of velvet antler. Online Journal of Veterinary Research. 10(1):20-25.

Cook N, Schaefer A, Warren L, Burwash L, Anderson M, Baron V. 2001. Adrenocortical and metabolic responses to ACTH injection in horses: an assessment by salivary cortisol and infrared thermography of the eye. Can J Anim Sc. 81.

Corujo LA, Kieson E, Schloesser T, Gloor PA. 2021. Emotion Recognition in Horses with Convolutional Neural Networks. Future Internet. 13(10):250.

Coulon M, Nowak R, Peyrat J, Chandèze H, Boissy A, Boivin X. 2015. Do Lambs Perceive Regular Human Stroking as Pleasant? Behavior and Heart Rate Variability Analyses. PLOS ONE. 10(2):e0118617.

d'Ingeo S, Quaranta A, Siniscalchi M, Stomp M, Coste C, Bagnard C, Hausberger M, Cousillas H. 2019. Horses associate individual human voices with the valence of past interactions: a behavioural and electrophysiological study. Scientific Reports. 9(1).

Dai F, Cogi NH, Heinzl EUL, Dalla Costa E, Canali E, Minero M. 2015. Validation of a fear test in sport horses using infrared thermography. Journal of Veterinary Behavior. 10(2):128-136.

Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC. 2014. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. PLOS ONE. 9(3):e92281.

Damasio AR, Grabowski TJ, Bechara A, Damasio H, Ponto LLB, Parvizi J, Hichwa RD. 2000. Subcortical and cortical brain activity during the feeling of self-generated emotions. Nature Neuroscience. 3(10):1049-1056.

Dawkins MS. 2006. Through animal eyes: What behaviour tells us. Applied Animal Behaviour Science. 100(1-2):4-10.

Dawkins MS. 2008. The science of animal suffering. Ethology. 114(10):937-945.

De Oliveira AR, Gozalo-Marcilla M, Ringer SK, Schauvliege S, Fonseca MW, Esteves Trindade PH, Prospero Puoli Filho JN, Luna SPL. 2021. Development and validation of the facial scale (FaceSed) to evaluate sedation in horses. PLOS ONE. 16(6):e0251909.

de Waal FBM. 2011. Year in Cognitive Neuroscience. Oxford: Blackwell Science Publ. PhD Positive emotions & welfare, What is an animal emotion?; p. 191-206.

DeBoer ML, Sheaffer CC, Grev AM, Catalano DN, Wells MS, Hathaway MR, Martinson KL. 2017. Yield, Nutritive Value, and Preference of Annual Warm-Season Grasses Grazed by Horses. Agronomy Journal. 109(5):2136-2148.

Demant H, Ladewig J, Balsby TJS, Dabelsteen T. 2011. The effect of frequency and duration of training sessions on acquisition and long-term memory in dogs. Applied Animal Behaviour Science. 133(3):228-234.

Des Roches ADB, Richard-Yris MA, Henry S, Ezzaouia M, Hausberger M. 2008. Laterality and emotions: Visual laterality in the domestic horse (Equus caballus) differs with objects' emotional value. Physiology & Behavior. 94(3):487-490.

Descovich KA, Wathan J, Leach MC, Buchanan-Smith HM, Flecknell P, Farningham D, Vick SJ. 2017. Facial expression: An under-utilised tool for the assessment of welfare in mammals. ALTEX. 34(3):409-429.

Dodman NH, Wrubel KM, Cottam N. 2008. Effect of a synthetic equine maternal pheromone during a controlled fear-eliciting situation: A critique on Falewee (2006). Applied Animal Behaviour Science. 109(1):85-87.

Dutton DW, Lashnits KJ, Wegner K. 2009. Managing severe hoof pain in a horse using multimodal analgesia and a modified composite pain score. Equine Veterinary Education. 21(1):37-43.

Ekman P. 2009. Darwin's contributions to our understanding of emotional expressions. Philosophical Transactions of the Royal Society B: Biological Sciences. 364(1535):3449-3451.

Esch T, Kim JW, Stefano GB. 2006. Neurobiological implications of eating healthy and its association with pleasure. Neuroendocrinology Letters. 27(1-2):21-33.

Falewee C, Gaultier E, Lafont C, Bougrat L, Pageat P. 2006. Effect of a synthetic equine maternal pheromone during a controlled fear-eliciting situation. Applied Animal Behaviour Science. 101(1-2):144-153.

Farmer K, Krueger K, Byrne RW. 2010. Visual laterality in the domestic horse (Equus caballus) interacting with humans. Animal Cognition. 13(2):229-238.

Feh C, de Mazieres J. 1993. Grooming at a preferred site reduces heart rate in horses. Animal Behaviour. 46(6):1191-1194.

Finlayson K, Lampe JF, Hintze S, Wurbel H, Melotti L. 2016. Facial Indicators of Positive Emotions in Rats. PLoS One. 11(11):e0166446.

Flecknell PA. 2010. Do mice have a pain face? Assessing changes in facial expression may enable us to assess pain in animals more accurately and more effectively.(NEWS AND VIEWS)(Report). Nature Methods. 7(6):437(2).

FLIR Systems. 2016. Infrared Camera Accuracy and Uncertainty in Plain Language. FLIR Systems; [accessed 2021 27/02/2021]. https://www.flir.com.au/discover/rd-science/infrared-camera-accuracy-and-uncertainty-in-plain-language/.

Forkman B, Boissy A, Meunier-Salauen MC, Canali E, Jones RB. 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiology & Behavior. 92(3):340-374.

Fraser D. 2008. Understanding animal welfare: the science in its cultural context. Kirkwood JK, Hubrecht RC ed. Oxford: Wiley-Blackwell.

Fraser D, Nicol CJ. 2018. Animal welfare. PhD Positive emotions & welfare, Preference and motivation research; p. 213-231.

Fredrickson BL. 2013. Advances in Experimental Social Psychology, Vol 47. PhD Positive emotions & welfare Non-PhD Groups, Positive Emotions Broaden and Build; p. 1-53.

Fureix C, Jego P, Sankey C, Hausberger M. 2009. How horses (Equus caballus) see the world: humans as significant "objects". Animal Cognition. 12(4):643-654.

Fureix C, Hausberger M, Seneque E, Morisset S, Baylac M, Cornette R, Biquand V, Deleporte P. 2011. Geometric morphometrics as a tool for improving the comparative study of behavioural postures. Naturwissenschaften. 98(7):583-592.

Garland EL, Fredrickson B, Kring AM, Johnson DP, Meyer PS, Penn DL. 2010. Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. Clinical Psychology Review. 30(7):849-864.

Garrett WS. 2017. A banner year for gut microbiota research. Nature Reviews Gastroenterology & Hepatology. 14(2):78-80.

Gehrke EK, Baldwin A, Schiltz PM. 2011. Heart Rate Variability in Horses Engaged in Equine-Assisted Activities. Journal of Equine Veterinary Science. 31(2):78-84.

Gibbs AE, Friend TH. 1999. Horse preference for orientation during transport and the effect of orientation on balancing ability. Applied Animal Behaviour Science. 63(1):1-9.

Giljov A, Karenina K. 2019. Differential roles of the right and left brain hemispheres in the social interactions of a free-ranging ungulate. Behavioural Processes. 168.

Gogoleva SS, Volodin IA, Volodina EV, Kharlamova AV, Trut LN. 2010. Sign and strength of emotional arousal: vocal correlates of positive and negative attitudes to humans in silver foxes (Vulpes vulpes). Behaviour. 147(13-14):1713-1736.

Goodwin D, Davidson HPB, Harris P. 2005. Selection and acceptance of flavours in concentrate diets for stabled horses. Applied Animal Behaviour Science. 95(3-4):223-232.

Grandin T. 2015. Improving animal welfare: a practical approach. Grandin T ed.

Grandin T, Deesing MJ. 2014 Elsevier. Behavioral Genetics and Animal Science; p. 1-40.

Grandin T, Odde KG, Schutz DN, Behrns LM. 1994. The reluctance of cattle to change a learned choice may confound preference tests. Applied Animal Behaviour Science. 39(1):21-28.

Green TC, Mellor DJ. 2011. Extending ideas about animal welfare assessment to include 'quality of life' and related concepts. New Zealand Veterinary Journal. 59(6):263-271.

Greiveldinger L, Veissier I, Boissy A. 2011. The ability of lambs to form expectations and the emotional consequences of a discrepancy from their expectations. Psychoneuroendocrinology. 36(6):806-815.

Grey AM, Sheaffer CC, DeBoer ML, Catalano DN, Martinson KL. 2017. Preference, Yield, and Forage Nutritive Value of Annual Grasses under Horse Grazing. Agronomy Journal. 109(4):1561-1572.

Griffin DR, Speck GB. 2004. New evidence of animal consciousness. Animal Cognition. 7(1):5-18.

Guesgen MJ. 2015. The social function of pain-related behaviour and novel techniques for the assessment of pain in lambs: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Zoology at Massey University, Turitea campus, Manawatu, New Zealand [Doctoral]. Massey University

Guesgen MJ, Bench CJ. 2017. What can kinematics tell us about the affective states of animals? Animal Welfare. 26(4):383-397.

Guesgen MJ, Beausoleil NJ, Minot EO, Stewart M, Stafford KJ. 2014. Social context and other factors influence the behavioural expression of pain by lambs. Applied Animal Behaviour Science. 159:41-49.

Guyton K, Houchen-Wise E, Peck E, Mayberry J. 2013. Equestrian Injury is Costly, Disabling, and Frequently Preventable: The Imperative for Improved Safety Awareness. The American Surgeon. 79(1):76-83.

Gygax L, Reefmann N, Pilheden T, Scholkmann F, Keeling L. 2015. Dog behavior but not frontal brain reaction changes in repeated positive interactions with a human: A non-invasive pilot study using functional near-infrared spectroscopy (fNIRS). Behavioural Brain Research. 281:172-176.

Hall C, Randle H, Pearson G, Preshaw L, Waran N. 2018. Assessing equine emotional state. Applied Animal Behaviour Science. 205:183-193.

Hall C, Huws N, White C, Taylor E, Owen H, McGreevy P. 2013. Assessment of ridden horse behavior. Journal of Veterinary Behavior-Clinical Applications and Research. 8(2):62-73.

Hall CA, Cassaday HJ, Derrington AM. 2003. The effect of stimulus height on visual discrimination in horses. Journal of Animal Science. 81(7):1715-1720.

Hamlin R, Klepinger W, Gilpin K, Smith C. 1972. Autonomic control of heart rate in the horse. American Journal of Physiology -- Legacy Content. 222(4):976-978.

Hamra JG, Kamerling SG, Wolfsheimer KJ, Bagwell CA. 1993. Diurnal-variation in plasma ir-beta-endorphin levels and experimental pain thresholds in the horse. Life Sciences. 53(2):121-129.

Hanggi EB. 2003. Discrimination learning based on relative size concepts in horses (Equus caballus). Applied Animal Behaviour Science. 83(3):201-213.

Hanggi EB. 2005 Lexington: American Association of Equine Practitioners (AAEP). PhD Positive emotions & welfare, The thinking horse: cognition and perception reviewed; p. 246-255.

Hanggi EB. 2010. Short-term Memory Testing in Domestic Horses: Experimental Design Plays a Role. Journal of Equine Veterinary Science. 30(11):617-623.

Hanggi EB, Ingersoll JF. 2009a. Stimulus discrimination by horses under scotopic conditions. Behavioural Processes. 82(1):45-50.

Hanggi EB, Ingersoll JF. 2009b. Long-term memory for categories and concepts in horses (Equus caballus). Animal Cognition. 12(3):451-462.

Hanggi EB, Ingersoll JF, Waggoner TL. 2007. Color vision in horses (Equus caballus): Deficiencies identified using a pseudoisochromatic plate test. Journal of Comparative Psychology. 121(1):65-72.

Hausberger M, Lesimple C, Henry S. 2021. Detecting Welfare in a Non-Verbal Species: Social/Cultural Biases and Difficulties in Horse Welfare Assessment. Animals. 11(8):2249.

Hausberger M, Roche H, Henry S, Visser EK. 2008. A review of the human-horse relationship. Applied Animal Behaviour Science. 109(1):1-24.

Hausberger M, Fureix C, Bourjade M, Wessel-Robert S, Richard-Yris MA. 2012. On the significance of adult play: what does social play tell us about adult horse welfare? Naturwissenschaften. 99(4):291-302.

Hellhammer DH, Wüst S, Kudielka BM. 2009. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology. 34(2):163-171.

Hemsworth PH, Rice M, Karlen MG, Calleja L, Barnett JL, Nash J, Coleman GJ. 2011. Human-animal interactions at abattoirs: Relationships between handling and animal stress in sheep and cattle. Applied Animal Behaviour Science. 135(1-2):24-33.

Henry S, Hemery D, Richard MA, Hausberger M. 2005. Human-mare relationships and behaviour of foals toward humans. Applied Animal Behaviour Science. 93(3-4):341-362.

Heuschele DJ, Catalano D, Martinson K, Wiersma J. 2018. Consumer Knowledge and Horse Preference for Different Colored Oats. Journal of Equine Veterinary Science. 71:6-12.

Hintze S, Schanz L. 2021. Using the Judgment Bias Task to Identify Behavioral Indicators of Affective State: Do Eye Wrinkles in Horses Reflect Mood? Frontiers in Veterinary Science. 8.

Hintze S, Smith S, Patt A, Bachmann I, Würbel H. 2016. Are eyes a mirror of the soul? What eye wrinkles reveal about a horse's emotional state. PLoS ONE. 11(10).

Hintze S, Murphy E, Bachmann I, Wemelsfelder F, Wurbel H. 2017. Qualitative Behaviour Assessment of horses exposed to short-term emotional treatments. Applied Animal Behaviour Science. 196:44-51.

Holcomb KE, Stull CL. 2016. Effect of time and weather on preference, frequency, and duration of shade use by horses. Journal of Animal Science. 94(4):1653-1661.

Holcomb KE, Tucker CB, Stull CL. 2014. Preference of domestic horses for shade in a hot, sunny environment. Journal of Animal Science. 92(4):1708-1717.

Houpt K, Marrow M, Seeliger M. 2000. A preliminary study of the effect of music on equine behavior. Journal of Equine Veterinary Science. 20(11):691-+.

Humphrey NK. 1972. 'Interest' and 'pleasure': two determinants of a monkey's visual preferences. Perception. 1(4):395-416.

Hunter L, Houpt KA. 1989. Bedding Material Preferences of Ponies. Journal of Animal Science. 67(8):1986.

Innes L, McBride S. 2008. Negative versus positive reinforcement: An evaluation of training strategies for rehabilitated horses. Applied Animal Behaviour Science. 112(3-4):357-368.

Inoue S, Yamamoto S, Ringhofer M, Mendonca RS, Hirata S. 2019. Lateral position preference in grazing feral horses. Ethology.

Iyasere OS, Durosaro SO, Oyeniran VJ, Daramola JO. 2022. Is an increase of glucocorticoid concentrations related to the degree of arousal or valence experienced by an animal to a stimulus? Domestic animal endocrinology. 81:106752-106752.

Janczarek I, Wilk I, Pietrzak S, Liss M, Tkaczyk S. 2018a. Taste Preferences of Horses in Relation to Their Breed and Sex. Journal of Equine Veterinary Science. 64:59-64.

Janczarek I, Stachurska A, Wilk I, Krakowski L, Przetacznik M, Zastrzezynska M, Kuna-Broniowska I. 2018b. Emotional excitability and behaviour of horses in response to stroking various regions of the body. Animal Science Journal. 89(11):1599-1608.

Jansen F, Van der Krogt J, Van Loon K, Avezzu V, Guarino M, Quanten S, Berckmans D. 2009. Online detection of an emotional response of a horse during physical activity. Veterinary Journal. 181(1):38-42.

Jansson A, Lindgren G, Solé M. 2021. An investigation into factors influencing basal eye temperature in the domestic horse (Equus caballus) when measured using infrared thermography in field conditions. Physiology & Behavior. 228.

Jensen P. 2015. Adding 'epi-' to behaviour genetics: implications for animal domestication. Journal of Experimental Biology. 218(1):32-40.

Jones AR, Smith A, Christey G. 2018. Equine-related injuries requiring hospitalisation in the Midland Region of New Zealand: a continuous five-year review. N Z Med J. 131(1483):50-58.

Jorgensen GHM, Mejdell CM, Boe KE. 2019. The effect of blankets on horse behaviour and preference for shelter in Nordic winter conditions. Applied Animal Behaviour Science. 218:104822-104822.

Jorgensen GHM, Aanensen L, Mejdell CM, Boe KE. 2016. Preference for shelter and additional heat in horses exposed to Nordic winter conditions. Equine Veterinary Journal. 48(6):720-726.

Jørgensen GHM, Liestøl SH-O, Bøe KE. 2011. Effects of enrichment items on activity and social interactions in domestic horses (Equus caballus). Applied Animal Behaviour Science. 129(2-4):100-110.

Kalbe C, Puppe B. 2010. Long-term cognitive enrichment affects opioid receptor expression in the amygdala of domestic pigs. Genes Brain and Behavior. 9(1):75-83.

Kato T, Ohmura H, Hiraga A, Wada S, Kuwahara M, Tsubone H. 2003. Changes in heart rate variability in horses during immersion in warm springwater. American Journal of Veterinary Research. 64(12):1482-1485.

Keeling LJ. 2019. Encyclopedia of Animal Behavior. PhD Positive emotions & welfare, Indicators of good welfare; p. 134-140.

Kelly KJ, McDuffee LA, Mears K. 2021. The Effect of Human–Horse Interactions on Equine Behaviour, Physiology, and Welfare: A Scoping Review. Animals. 11(10):2782.

Kirkden RD, Pajor EA. 2006. Using preference, motivation and aversion tests to ask scientific questions about animals' feelings. Applied Animal Behaviour Science. 100(1–2):29-47.

Kirkwood JK, Hubrecht R. 2001. Animal consciousness, cognition and welfare. Animal Welfare. 10:S5-S17.

Knutson B, Burgdorf J, Panksepp J. 2002. Ultrasonic Vocalizations as indices of affective states in rats. Psychological Bulletin. 128(6):961-977.

Kok BE, Fredrickson BL. 2010. Upward spirals of the heart: Autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. Biological Psychology. 85(3):432-436.

Konig von Borstel U, Duncan IJH, Shoveller AK, Merkies K, Keeling LJ, Millman ST. 2009. Impact of riding in a coercively obtained Rollkur posture on welfare and fear of performance horses. Applied Animal Behaviour Science. 116(2-4):228-236.

König von Borstel U, Keil J. 2012. Horses' behavior and heart rate in a preference test for shorter and longer riding bouts. Journal of Veterinary Behavior: Clinical Applications and Research. 7(6):362-374.

Koopman M, El Aidy S, Consortium MI. 2017. Depressed gut? The microbiota-diet-inflammation trialogue in depression. Current Opinion in Psychiatry. 30(5):369-377.

Korte SM, Olivier B, Koolhaas JM. 2007. A new animal welfare concept based on allostasis. Physiology & Behavior. 92(3):422-428.

Kowalik S, Janczarek I, Kędzierski W, Stachurska A, Wilk I. 2017. The effect of relaxing massage on heart rate and heart rate variability in purebred Arabian racehorses. Animal Science Journal. 88(4):669-677.

Krawczel PD, Friend TH, Johnson R. 2006. A note on the preference of naive horses for different water bowls. Applied Animal Behaviour Science. 100(3-4):309-313.

Kremer L, Klein Holkenborg SEJ, Reimert I, Bolhuis JE, Webb LE. 2020. The nuts and bolts of animal emotion. Neuroscience & Biobehavioral Reviews. 113:273-286.

Krueger K, Esch L, Farmer K, Marr I. 2021. Basic needs in horses?—a literature review. Animals. 11(6).

Krugmann K, Warnken F, Krieter J, Czycholl I. 2019. Are behavioral tests capable of measuring positive affective states in growing pigs? Animals. 9(5).

Kuwahara M, Hashimoto S, Ishii K, Yagi Y, Hada T, Hiraga A, Kai M, Kubo K, Oki H, Tsubone H et al. 1996. Assessment of autonomic nervous function by power spectral analysis of heart rate variability in the horse. Journal of the Autonomic Nervous System. 60(1-2):43-48.

Kuznetsova M, Wilson C, Hannan AJ, Renoir T. 2020. How the enriched get richer? Experience-dependent modulation of microRNAs and the therapeutic effects of environmental enrichment. Pharmacology Biochemistry and Behavior. 195:172940.

Laerd Statistics. 2015. One-way repeated measures ANOVA using SPSS Statistics. [accessed 18Feb2022]. https://statistics.laerd.com/premium/spss/owrma/one-way-repeated-measures-anova-in-spss-17.php.

LaFollette MR, O'Haire ME, Cloutier S, Gaskill BN. 2018. Practical rat tickling: Determining an efficient and effective dosage of heterospecific play. Applied Animal Behaviour Science. 208:82-91.

Laister S, Stockinger B, Regner AM, Zenger K, Knierim U, Winckler C. 2011. Social licking in dairy cattle-Effects on heart rate in performers and receivers. Applied Animal Behaviour Science. 130(3-4):81-90.

Lanata A, Guidi A, Baragli P, Valenza G, Scilingo EP. 2015. A Novel Algorithm for Movement Artifact Removal in ECG Signals Acquired from Wearable Systems Applied to Horses. PLOS ONE. 10(10):e0140783.

Lanata A, Nardelli M, Valenza G, Baragli P, Draniello B, Alterisio A, Scandurra A, Semin GR, Scilingo EP. 2018. A Case for the Interspecies Transfer of Emotions: A Preliminary Investigation on How Humans Odors Modify Reactions of the Autonomic Nervous System in Horses. Proceedings of 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018.

Lange A, Franzmayr S, Wisenöcker V, Futschik A, Waiblinger S, Lürzel S. 2020. Effects of different stroking styles on behaviour and cardiac parameters in heifers. Animals. 10(3).

Langford DJ, Bailey AL, Chanda ML, Clarke SE, Drummond TE, Echols S, Glick S, Ingrao J, Klassen-Ross T, LaCroix-Fralish ML et al. 2010. Coding of facial expressions of pain in the laboratory mouse. Nature Methods. 7(6):447-U52.

Lansade L, Bouissou MF. 2008a. Reactivity to humans: A temperament trait of horses which is stable across time and situations. Applied Animal Behaviour Science. 114(3-4):492-508.

Lansade L, Simon F. 2010. Horses' learning performances are under the influence of several temperamental dimensions. Applied Animal Behaviour Science. 125(1-2):30-37.

Lansade L, Bouissou MF, Erhard HW. 2008b. Reactivity to isolation and association with conspecifics: A temperament trait stable across time and situations. Applied Animal Behaviour Science. 109(2-4):355-373.

Lansade L, Bouissou MF, Erhard HW. 2008c. Fearfulness in horses: A temperament trait stable across time and situations. Applied Animal Behaviour Science. 115(3-4):182-200.

Lansade L, Bonneau C, Parias C, Biau S. 2019. Horse's emotional state and rider safety during grooming practices, a field study. Applied Animal Behaviour Science. 217:43-47.

Lansade L, Nowak R, Lainé AL, Leterrier C, Bonneau C, Parias C, Bertin A. 2018. Facial expression and oxytocin as possible markers of positive emotions in horses. Scientific Reports. 8(1):14680.

Lawrence AB, Vigors B, Sandoe P. 2019. What is so positive about positive animal welfare? A critical review of the literature. Animals. 9(10).

Leach MC, Coulter CA, Richardson CA, Flecknell PA. 2011. Are We Looking in the Wrong Place? Implications for Behavioural-Based Pain Assessment in Rabbits (Oryctolagus cuniculi) and Beyond? Plos One. 6(3).

Lebelt D, Zanella AJ, Unshelm J. 1998. Physiological correlates associated with cribbing behaviour in horses: changes in thermal threshold, heart rate, plasma β-endorphin and serotonin. Equine Veterinary Journal. 30(S27):21-27.

Lee J, Floyd T, Erb H, Houpt K. 2011. Preference and demand for exercise in stabled horses. Applied Animal Behaviour Science. 130(3–4):91-100.

Lee S-H, Yoon S-H, Jung Y, Kim N, Min U, Chun J, Choi I. 2020. Emotional well-being and gut microbiome profiles by enterotype. Scientific Reports. 10(1).

Leliveld LMC, 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.

Lenoir A, Trachsel DS, Younes M, Barrey E, Robert C. 2017. Agreement between Electrocardiogram and Heart Rate Meter Is Low for the Measurement of Heart Rate Variability during Exercise in Young Endurance Horses. Frontiers in Veterinary Science. 4.

Lerche P. 2009. Assessment and treatment of pain in horses. Equine Veterinary Education. 21:44-45.

LeScolan N, Hausberger M, Wolff A. 1997. Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. Behavioural Processes. 41(3):257-266.

Lesimple C. 2020. Indicators of Horse Welfare: State-of-the-Art. Animals. 10(2).

Levitis DA, Lidicker WZ, Freund G. 2009. Behavioural biologists do not agree on what constitutes behaviour. Animal Behaviour. 78(1):103-110.

Lezama-García K, Orihuela A, Olmos-Hernández A, Reyes-Long S, Mota-Rojas D. 2019. Facial expressions and emotions in domestic animals. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources. 14.

Lie M, Newberry RC. 2018. Horse communication: what does non-nutritive chewing mean? In: McDonnell S, Padalino B, Baragli P, editor(s). Proceedings of the 14th International Equitation Science Conference; Sep 21-24; Rome, Italy. p. 49.

Loeser J. 2011. IASP Taxonomy. [accessed 2012 21 March]. http://www.iasp-pain.org/AM/Template.cfm?Section=Pain_Defi...isplay.cfm&ContentID=1728.

Love EJ. 2012. Equine Surgery (Fourth Edition). Saint Louis: W.B. Saunders. PhD Positive emotions & welfare, Chapter 23 - Equine Pain Management; p. 263-270.

Lundblad J. 2018. Changes in facial expressions during short term emotional stress as described by a Facial Action Coding System in horses. Uppsala: Swedish University of Agricultural Sciences

Lv J, Li J, Wang C, Zhao P, Bi Y, Zhang X, Yi R, Li X, Bao J. 2018. Positive or negative emotion induced by feeding success or failure can affect behaviors, heart rate and immunity of suckling calves. Physiology and Behavior. 196:185-189.

Macchitella L, Stegagno T, Giaconella R, Polizzi di Sorrentino E, Schino G, Addessi E. 2017. Oxytocin improves the ability of dogs to follow informative pointing: a neuroemotional hypothesis. Rendiconti Lincei. 28(1):105-115.

Marcet-Rius M, Kalonji G, Cozzi A, Bienboire-Frosini C, Monneret P, Kowalczyk I, Teruel E, Codecasa E, Pageat P. 2019. Effects of straw provision, as environmental enrichment, on behavioural indicators of welfare and emotions in pigs reared in an experimental system. Livestock Science. 221:89-94.

Marcet Rius M, Cozzi A, Bienboire-Frosini C, Teruel E, Chabaud C, Monneret P, Leclercq J, Lafont-Lecuelle C, Pageat P. 2018. Providing straw to allow exploratory behaviour in a pig experimental system does not modify putative indicators of positive welfare: peripheral oxytocin and serotonin. Animal. 12(10):2138-2146.

Martin P, Bateson P. 2007. Measuring behaviour: an introductory guide. 3rd ed. Cambridge: Cambridge University Press.

Martin TI, Zentall TR, Lawrence L. 2006. Simple discrimination reversals in the domestic horse (Equus caballus): Effect of discriminative stimulus modality on learning to learn. Applied Animal Behaviour Science. 101(3-4):328-338.

Mauss D, Jarczok MN, Fischer JE. 2016. The streamlined Allostatic Load Index: a replication of study results. Stress. 19(6):553-558.

Mauss IB, Robinson MD. 2009. Measures of emotion: A review. Cognition & Emotion. 23(2):209-237.

McBride SD, Hemmings A, Robinson K. 2004. A preliminary study on the effect of massage to reduce stress in the horse. Journal of Equine Veterinary Science. 24(2):76-81.

McBride SD, Parker MO, Roberts K, Hemmings A. 2017. Applied neurophysiology of the horse; implications for training, husbandry and welfare. Applied Animal Behaviour Science. 190:90-101.

McDonnell S. 2003. Practical field guide to horse behavior: the equid ethogram. McDonnell S ed. Lexington: The Blood-Horse, Inc.

McDonnell SM, Haviland JCS. 1995. Agonistic ethogram of the equid bachelor band. Applied Animal Behaviour Science. 43(3):147-188.

McDonnell SM, Poulin A. 2002. Equid play ethogram. Applied Animal Behaviour Science. 78(2-4):263-290.

McDowell LJ, Wells DL, Hepper PG, Dempster M. 2016. Lateral bias and temperament in the domestic cat (Felis silvestris). Journal of Comparative Psychology. 130(4):313-320.

McGreevy P. 2004. Equine behavior : a guide for veterinarians and equine scientists. Edinburgh: Saunders.

McGreevy P. 2012. Equine Behavior: A Guide for Veterinarians and Equine Scientists. 2nd ed. United Kingdom: Elsevier Health Sciences.

McGreevy P, Warren-Smith A, Guisard Y. 2012. The effect of double bridles and jaw-clamping crank nosebands on temperature of eyes and facial skin of horses. Journal of Veterinary Behavior: Clinical Applications and Research. 7(3):142-148.

McGreevy P, Christensen JW, Konig von Borstel U, McLean A. 2018. Equitation Science: Wiley.

Mejdell CM, Buvik T, Jorgensen GHM, Boe KE. 2016. Horses can learn to use symbols to communicate their preferences. Applied Animal Behaviour Science. 184:66-73.

Mejdell CM, Jorgensen GHM, Buvik T, Torp T, Boe KE. 2019. The effect of weather conditions on the preference in horses for wearing blankets. Applied Animal Behaviour Science. 212:52-57.

Mellor DJ. 2015. Enhancing animal welfare by creating opportunities for positive affective engagement. New Zealand Veterinary Journal. 63(1):3-8.

Mellor DJ. 2016. Updating animal welfare thinking: Moving beyond the “Five Freedoms” towards “A life worth living”. Animals. 6(3):21.

Mellor DJ, Beausoleil NJ. 2015. Extending the 'Five Domains' model for animal welfare assessment to incorporate positive welfare states. Animal Welfare. 24(3):241-253.

Mendl M, Paul ES. 2004. Consciousness, emotion and animal welfare: insights from cognitive science. Animal Welfare. 13:S17-S25.

Mendl M, Burman OHP, Paul ES. 2010. An integrative and functional framework for the study of animal emotion and mood. Proceedings of the Royal Society B: Biological Sciences. 277(1696):2895-2904.

Merkies, Ready, Farkas, Hodder. 2019. Eye Blink Rates and Eyelid Twitches as a Non-Invasive Measure of Stress in the Domestic Horse. Animals. 9(8):562.

Miller LJ, Vicino GA, Sheftel J, Lauderdale LK. 2020. Behavioral Diversity as a Potential Indicator of Positive Animal Welfare. Animals. 10(7):17.

Mills DS, Nankervis KJ. 1999. Equine behaviour : principles and practice. Malden, Mass. ; Oxford: Blackwell Science.

Mills DS, Eckley S, Cooper JJ. 2000. Thoroughbred bedding preferences, associated behaviour differences and their implications for equine welfare. Animal Science. 70:95-106.

Minero M, Tosi MV, Canali E, Wemelsfelder F. 2009. Quantitative and qualitative assessment of the response of foals to the presence of an unfamiliar human. Applied Animal Behaviour Science. 116(1):74-81.

Mitsui S, Yamamoto M, Nagasawa M, Mogi K, Kikusui T, Ohtani N, Ohta M. 2011. Urinary oxytocin as a noninvasive biomarker of positive emotion in dogs. Hormones and Behavior. 60(3):239-243.

Mogil JS, Pang DSJ, Silva Dutra GG, Chambers CT. 2020. The development and use of facial grimace scales for pain measurement in animals. Neuroscience & Biobehavioral Reviews. 116:480-493.

Muehlemann T, Reefmann N, Wechsler B, Wolf M, Gygax L. 2011. In vivo functional near-infrared spectroscopy measures mood-modulated cerebral responses to a positive emotional stimulus in sheep. Neuroimage. 54(2):1625-1633.

Muller CE, Uden P. 2007. Preference of horses for grass conserved as hay, haylage or silage. Animal Feed Science and Technology. 132(1-2):66-78.

Napolitano F, De Rosa G, Braghieri A, Grasso F, Bordi A, Wemelsfelder F. 2008. The qualitative assessment of responsiveness to environmental challenge in horses and ponies. Applied Animal Behaviour Science. 109(2-4):342-354.

Neethirajan S, Reimert I, Kemp B. 2021. Measuring Farm Animal Emotions—Sensor-Based Approaches. Sensors. 21(2):553.

New Zealand Thoroughbred Racing 2019. Thoroughbred Welfare Assessment Guidelines. New Zealand Thoroughbred Racing.

Normando S, Trevisan C, Bonetti O, Bono G. 2007. A note on heart rate response to massage in stereotyping and non stereotyping horses. Journal of Animal and Veterinary Advances. 6(1):101-104.

Normando S, Haverbeke A, Meers L, Odberg FO, Talegon MI, Bono G. 2003. Effect of manual imitation of grooming on riding horses' heart rate in different environmental situations. Veterinary Research Communications. 27:615-617.

Northey G. 2003. Equestrian injuries in New Zealand, 1993-2001: knowledge and experience. The New Zealand Medical Journal (Online). 116(1182).

O'Connor CI, Ayers C, Nielsen BD. 2004. Preference of complete pelleted feeds with varying concentrate to forage ratios in two ages of horses. Journal of Animal and Veterinary Advances. 3(3):179-182.

Ochoa-Repáraz J, Kasper LH. 2016. The Second Brain: Is the Gut Microbiota a Link Between Obesity and Central Nervous System Disorders? Current Obesity Reports. 5(1):51-64.

Odendaal JS, Meintjes RA. 2003. Neurophysiological correlates of affiliative behaviour between humans and dogs. Vet J. 165(3):296-301.

Overall K. 2013. Manual of Clinical Behavioral Medicine for Dogs and Cats. United Kingdom: Elsevier.

Pageat P. 2008. Response to Dodman, Wrubel & Cottam: Effect of a synthetic equine maternal pheromone during a controlled fear-eliciting situation: A critique on Falewee (January 2006). Applied Animal Behaviour Science. 109(1):88-93.

Panduro A, Rivera-Iñiguez I, Sepulveda-Villegas M, Roman S. 2017. Genes, emotions and gut microbiota: The next frontier for the gastroenterologist. World Journal of Gastroenterology. 23(17):3030.

Panksepp J. 2005. Affective consciousness: Core emotional feelings in animals and humans. Consciousness and Cognition. 14(1):30-80.

Panksepp J. 2010. Affective consciousness in animals: perspectives on dimensional and primary process emotion approaches. Proceedings of the Royal Society B-Biological Sciences. 277(1696):2905-2907.

Panksepp J, Lahvis GP. 2011. Rodent empathy and affective neuroscience. Neuroscience and Biobehavioral Reviews. 35(9):1864-1875.

Panksepp J, Lane RD, Solms M, Smith R. 2017. Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. Neuroscience & Biobehavioral Reviews. 76:187-215.

Paul ES, Mendl MT. 2018. Animal emotion: Descriptive and prescriptive definitions and their implications for a comparative perspective. Applied Animal Behaviour Science. 205:202-209.

Paul ES, Harding EJ, Mendl M. 2005. Measuring emotional processes in animals: the utility of a cognitive approach. Neuroscience & Biobehavioral Reviews. 29(3):469-491.

Paul ES, Sher S, Tamietto M, Winkielman P, Mendl MT. 2020. Towards a comparative science of emotion: Affect and consciousness in humans and animals. Neuroscience & Biobehavioral Reviews. 108:749-770.

Petersson M, Uvnäs-Moberg K, Nilsson A, Gustafson L-L, Hydbring-Sandberg E, Handlin L. 2017. Oxytocin and Cortisol Levels in Dog Owners and Their Dogs Are Associated with Behavioral Patterns: An Exploratory Study. Frontiers in Psychology. 8.

Petrie A, Watson P. 2013. Statistics for veterinary and animal science. 3rd ed. West Sussex, UK: John Wiley & Sons, Ltd.

Pfister JA, Lee ST, Arnett DP, Panter KE. 2017. Preference by horses for bedding pellets made from switchgrass (Panicum virgatum) straw. Professional Animal Scientist. 33(3):349-356.

Proctor HS, Carder G. 2015. Nasal temperatures in dairy cows are influenced by positive emotional state. Physiology & Behavior. 138:340-344.

Quaranta A, Siniscalchi M, Vallortigara G. 2007. Asymmetric tail-wagging responses by dogs to different emotive stimuli. Current Biology. 17(6):R199-R201.

Randall L, Rogers CW, Hoskin SO, Morel PC, Swainson NM. 2014. Preference for different pasture grasses by horses in New Zealand. Proceedings of the New Zealand Society of Animal Production. 74:79-84.

Rault J-L, Van Den Munkhof M, Buisman-Pijlman FTA. 2017. Oxytocin as an Indicator of Psychological and Social Well-Being in Domesticated Animals: A Critical Review. Frontiers in Psychology. 8.

Redaelli V, Luzi F, Mazzola S, Bariffi G, Zappaterra M, Nanni Costa L, Padalino B. 2019. The Use of Infrared Thermography (IRT) as Stress Indicator in Horses Trained for Endurance: A Pilot Study. Animals. 9(3):84.

Reefmann N, Wechsler B, Gygax L. 2009a. Behavioural and physiological assessment of positive and negative emotion in sheep. Animal Behaviour. 78(3):651-659.

Reefmann N, Kaszas FB, Wechsler B, Gygax L. 2009b. Ear and tail postures as indicators of emotional valence in sheep. Applied Animal Behaviour Science. 118(3-4):199-207.

Reefmann N, Kaszas FB, Wechsler B, Gygax L. 2009c. Physiological expression of emotional reactions in sheep. Physiology & Behavior. 98(1-2):235-241.

Rekant SI, Lyons MA, Pacheco JM, Arzt J, Rodriguez LL. 2016. Veterinary applications of infrared thermography. American Journal of Veterinary Research. 77(1):98-107.

Rezakhani A, Goodarzi M, Mokhber-Dezfully MR. 2005. Sinus bradycardia in a horse. Pakistan Veterinary Journal. 25(1):40-42.

Riemer S, Thompson H, Burman OHP. 2018. Behavioural responses to unexpected changes in reward quality. Scientific Reports. 8(1):16652.

Rietmann TR, Stauffacher M, Bernasconi P, Auer JA, Weishaupt MA. 2004a. The association between heart rate, heart rate variability, endocrine and behavioural pain measures in horses suffering from laminitis. Journal of Veterinary Medicine Series a-Physiology Pathology Clinical Medicine. 51(5):218-225.

Rietmann TR, Stuart AEA, Bernasconi P, Stauffacher M, Auer JA, Weishaupt MA. 2004b. Assessment of mental stress in warmblood horses: heart rate variability in comparison to heart rate and selected behavioural parameters. Applied Animal Behaviour Science. 88(1-2):121-136.

Rolls ET. 2000. Precis of The brain and emotion. Behav Brain Sci. 23(2):177-234.

Russell JA. 1980. A circumplex model of affect. Journal of Personality and Social Psychology. 39(6):1161-1178.

Sabry SM, Ola A, Nadia B, Sanaa R, Hend G. 2020. Prediction of Health Risk and Estimation of Associated Variables with Work Stress using Allostatic Load Index. Biomedical and Pharmacology Journal. 13:979-987.

Sandem AI, Braastad BO, Bøe KE. 2002. Eye white may indicate emotional state on a frustration–contentedness axis in dairy cows. Applied Animal Behaviour Science. 79(1):1-10.

Sandem AI, Janczak AM, Braastad BO. 2004. A short note on effects of exposure to a novel stimulus (umbrella) on behaviour and percentage of eye-white in cows. Applied Animal Behaviour Science. 89(3-4):309-314.

Sandem AI, Janczak AM, Salte R, Braastad BO. 2006. The use of diazepam as a pharmacological validation of eye white as an indicator of emotional state in dairy cows. Applied Animal Behaviour Science. 96(3-4):177-183.

Sankey C, Henry S, Gorecka-Bruzda A, Richard-Yris MA, Hausberger M. 2010a. The Way to a Man's Heart Is through His Stomach: What about Horses? Plos One. 5(11).

Sankey C, Richard-Yris MA, Leroy H, Henry S, Hausberger M. 2010b. Positive interactions lead to lasting positive memories in horses, Equus caballus. Animal Behaviour. 79(4):869-875.

Sankey C, Richard-Yris MA, Henry S, Fureix C, Nassur F, Hausberger M. 2010c. Reinforcement as a mediator of the perception of humans by horses (Equus caballus). Animal Cognition. 13(5):753-764.

Sappington BKF, McCall CA, Coleman DA, Kuhlers DL, Lishak RS. 1997. A preliminary study of the relationship between discrimination reversal learning and performance tasks in yearling and 2-year-old horses. Applied Animal Behaviour Science. 53(3):157-166.

Schilder MBH, van Hooff JARAM, van Geer-Plesman CJ, Wensing JB. 1984. A Quantitative Analysis of Facial Expression in the Plains Zebra. Zeitschrift für Tierpsychologie. 66(1):11-32.

Schmidt A, Hödl S, Möstl E, Aurich J, Müller J, Aurich C. 2010. Cortisol release, heart rate, and heart rate variability in transport-naive horses during repeated road transport. Domestic Animal Endocrinology. 39(3):205-213.

Schmied C, Waiblinger S, Scharl T, Leisch F, Boivin X. 2008. Stroking of different body regions by a human: Effects on behaviour and heart rate of dairy cows. Applied Animal Behaviour Science. 109(1):25-38.

Schrimpf A, Single M-S, Nawroth C. 2020. Social Referencing in the Domestic Horse. Animals. 10(1).

Scopa C, Greco A, Contalbrigo L, Fratini E, Lanatà A, Scilingo EP, Baragli P. 2020. Inside the Interaction: Contact With Familiar Humans Modulates Heart Rate Variability in Horses. Frontiers in Veterinary Science. 7.

Seaman SC, Davidson HPB, Waran NK. 2002. How reliable is temperament assessment in the domestic horse (Equus caballus)? Applied Animal Behaviour Science. 78(2-4):175-191.

Semin GR, Scandurra A, Baragli P, Lanata A, D'Aniello B. 2019. Inter- and Intra-Species Communication of Emotion: Chemosignals as the Neglected Medium. Animals. 9(11).

Serrapica M, Boivin X, Coulon M, Braghieri A, Napolitano F. 2017. Positive perception of human stroking by lambs: Qualitative behaviour assessment confirms previous interpretation of quantitative data. Applied Animal Behaviour Science. 187:31-37.

Shammas MA. 2011. Telomeres, lifestyle, cancer, and aging. Current Opinion in Clinical Nutrition and Metabolic Care. 14(1):28-34.

Shimada M, Suzuki N. 2020. The Contribution of Mutual Grooming to Affiliative Relationships in a Feral Misaki Horse Herd. Animals. 10(9):1564.

Siniscalchi M, Padalino B, Lusito R, Quaranta A. 2014. Is the left forelimb preference indicative of a stressful situation in horses? Behavioural Processes. 107:61-67.

Siniscalchi M, D'Ingeo S, Fornelli S, Quaranta A. 2018. Lateralized behavior and cardiac activity of dogs in response to human emotional vocalizations. Scientific Reports. 8(1).

Smith AV, Wilson C, McComb K, Proops L. 2018. Domestic horses (Equus caballus) prefer to approach humans displaying a submissive body posture rather than a dominant body posture. Animal Cognition. 21(2):307-312.

Smith AV, Proops L, Grounds K, Wathan J, McComb K. 2016. Functionally relevant responses to human facial expressions of emotion in the domestic horse (Equus caballus). Biology Letters. 12(3).

Smith BL, Jones JH, Carlson GP, Pascoe JR. 1994. Body position and direction preferences in horses during road transport. Equine Veterinary Journal. 26(5):374-377.

Smith JD. 2009. The study of animal metacognition. Trends in Cognitive Sciences. 13(9):389-396.

Smith PM, Ferguson AV. 2008. Neurophysiology of hunger and satiety. Developmental Disabilities Research Reviews. 14(2):96-104.

Snowdon CT, Teie D, Savage M. 2015. Cats prefer species-appropriate music. Applied Animal Behaviour Science. 166(0):106-111.

Søndergaard E, Jensen MB, Nicol CJ. 2011. Motivation for social contact in horses measured by operant conditioning. Applied Animal Behaviour Science. 132(3-4):131-137.

Sotocinal SG, Sorge RE, Zaloum A, Tuttle AH, Martin LJ, Wieskopf JS, Mapplebeck JCS, Wei P, Zhan S, Zhang SR et al. 2011. The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. Molecular Pain. 7.

Spiesberger K, Lürzel S, Patzl M, Futschik A, Waiblinger S. 2019. The effects of play behavior, feeding, and time of day on salivary concentrations of sIgA in calves. Animals. 9(9).

Spinka M, Duncan IJH, Widowski TM. 1998. Do domestic pigs prefer short-term to medium-term confinement? Applied Animal Behaviour Science. 58(3-4):221-232.

Stachurska A, Sniadek J, Janczarek I. 2017. First reaction of the heart rate in horses to different genres of music played in the stable. Medycyna Weterynaryjna-Veterinary Medicine-Science and Practice. 73(8):500-504.

Stachurska A, Janczarek I, Wilk I, Kedzierski W. 2015. Does Music Influence Emotional State in Race Horses? Journal of Equine Veterinary Science. 35(8):650-656.

Steptoe A, Wardle J, Marmot M. 2005. Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes. Proceedings of the National Academy of Sciences of the United States of America. 102(18):6508-6512.

Stewart M, Webster JR, Schaefer AL, Cook NJ, Scott SL. 2005. Infrared thermography as a non-invasive tool to study animal welfare. Animal Welfare. 14(4):319-325.

Stewart M, Stafford KJ, Dowling SK, Schaefer AL, Webster JR. 2008a. Eye temperature and heart rate variability of calves disbudded with or without local anaesthetic. Physiology & Behavior. 93(4-5):789-797.

Stewart M, Webster JR, Stafford KJ, Schaefer AL, Verkerk GA. 2010a. Technical note: Effects of an epinephrine infusion on eye temperature and heart rate variability in bull calves. Journal of Dairy Science. 93(11):5252-5257.

Stewart M, Verkerk GA, Stafford KJ, Schaefer AL, Webster JR. 2010b. Noninvasive assessment of autonomic activity for evaluation of pain in calves, using surgical castration as a model. Journal of Dairy Science. 93(8):3602-3609.

Stewart M, Stratton RB, Beausoleil NJ, Stafford KJ, Worth GM, Waran NK. 2011. Assessment of positive emotions in horses: Implications for welfare and performance. Journal of Veterinary Behavior: Clinical Applications and Research. 6(5):296.

Stewart M, Schaefer A, Hale DB, Colyn J, Cook NJ, Stafford KJ, Webster JR. 2008b. Infrared thermography as a non-invasive method for detecting fear-related responses of cattle to handling procedures. Animal Welfare. 17(4):387-393.

Stomp M, Masson A, Henry S, Hausberger M, Lesimple C. 2020. Could snorts inform us on how horses perceive riding? Behavioural Processes. 172.

Stomp M, D'Ingeo S, Henry S, Cousillas H, Hausberger M. 2021. Brain activity reflects (chronic) welfare state: Evidence from individual electroencephalography profiles in an animal model. Applied Animal Behaviour Science. 236:105271.

Stomp M, Leroux M, Cellier M, Henry S, Lemasson A, Hausberger M. 2018. An unexpected acoustic indicator of positive emotions in horses. PLoS ONE. 13(7).

Stratton R, Cogger N, Beausoleil N, Waran N, Stafford K, Stewart M. 2014. Technical paper: Indicators of Good Welfare in Horses. Wellington, New Zealand, New Zealand Ministry for Primary Industries.

Stubsjoen SM, Flo AS, Moe RO, Janczak AM, Skjerve E, Valle PS, Zanella AJ. 2009. Exploring non-invasive methods to assess pain in sheep. Physiology & Behavior. 98(5):640-648.

Susskind JM, Anderson AK. 2008. Facial expression form and function. Communicative & integrative biology. 1(2):148-149.

Taborsky M. 2010. Sample Size in the Study of Behaviour. Ethology. 116(3):185-202.

Tamioso PR, Maiolino Molento CF, Boivin X, Chandèze H, Andanson S, Delval É, Hazard D, da Silva GP, Taconeli CA, Boissy A. 2018. Inducing positive emotions: Behavioural and cardiac responses to human and brushing in ewes selected for high vs low social reactivity. Applied Animal Behaviour Science. 208:56-65.

Taylor PM, Pascoe PJ, Mama KR. 2002. Diagnosing and treating pain in the horse - Where are we today? Veterinary Clinics of North America-Equine Practice. 18(1):1-+.

Thorbergson ZW, Nielsen SG, Beaulieu RJ, Doyle RE. 2016. Physiological and Behavioral Responses of Horses to Wither Scratching and Patting the Neck When Under Saddle. Journal of Applied Animal Welfare Science. 19(3):245-259.

Trösch M, Pellon S, Cuzol F, Parias C, Nowak R, Calandreau L, Lansade L. 2020. Horses feel emotions when they watch positive and negative horse–human interactions in a video and transpose what they saw to real life. Animal Cognition. 23(4):643-653.

Tsai V, Chen N. 2020. A Gut Feeling. Grey Matters(17).

Usui H, Nishida Y. 2017. The very low-frequency band of heart rate variability represents the slow recovery component after a mental stress task. PLOS ONE. 12(8):e0182611.

van den Berg M, Giagos V, Lee C, Brown WY, Cawdell-Smith AJ, Hinch GN. 2016. The influence of odour, taste and nutrients on feeding behaviour and food preferences in horses. Applied Animal Behaviour Science. 184:41-50.

van Dierendonck MC, Sigurjónsdóttir H, Colenbrander B, Thorhallsdottir A. 2004. Differences in social behaviour between late pregnant, post-partum and barren mares in a herd of Icelandic horses. Applied Animal Behaviour Science. 89.

van Dierendonck MV, Burden F, Rickards KJ, van Loon JV. 2020. Monitoring Acute Pain in Donkeys with the Equine Utrecht University Scale for Donkeys Composite Pain Assessment (EQUUS-DONKEY-COMPASS) and the Equine Utrecht University Scale for Donkey Facial Assessment of Pain (EQUUS-DONKEY-FAP). Animals. 10(354).

van Loon J, Back W, Hellebrekers LJ, van Weeren PR. 2010. Application of a Composite Pain Scale to Objectively Monitor Horses with Somatic and Visceral Pain under Hospital Conditions. Journal of Equine Veterinary Science. 30(11):641-649.

van Oostrom H, Stienen PJ, van den Bos R, de Groot HNM. 2007. Somatosensory-evoked potentials indicate increased unpleasantness of noxious stimuli in response to increasing stimulus intensities in the rat. Brain Research Bulletin. 71(4):404-409.

Visser EK, Van Reenen CG, Schilder MBH, Barneveld A, Blokhuis HJ. 2003. Learning performances in young horses using two different learning tests. Applied Animal Behaviour Science. 80(4):311-326.

Visser EK, van Reenen CG, van der Werf JTN, Schilder MBH, Knaap JH, Barneveld A, Blokhuis HJ. 2002. Heart rate and heart rate variability during a novel object test and a handling test in young horses. Physiology & Behavior. 76(2):289-296.

von Borell E, Langbein J, Despres G, Hansen S, Leterrier C, Marchant-Forde J, Marchant-Forde R, Minero M, Mohr E, Prunier A et al. 2007. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - A review. Physiology & Behavior. 92(3):293-316.

Wager T, Barrett L, Bliss-Moreau E, Lindquist K, Duncan S, Kober H, Joseph J, Davidson M, Mize J. 2008. Handbook of emotions (3rd ed.). New York: Guilford Press. PhD Positive emotions & welfare, The neuroimaging of emotion; p. 249-271.

Walker J, Dale A, Waran N, Clarke N, Farnworth M, Wemelsfelder F. 2010. The assessment of emotional expression in dogs using a Free Choice Profiling methodology. Animal Welfare. 19(1):75-84.

Waller BM, Julle-Daniere E, Micheletta J. 2020. Measuring the evolution of facial 'expression' using multi-species FACS. Neuroscience & Biobehavioral Reviews. 113:1-11.

Wang H-M, Huang S-C. 2012. SDNN/RMSSD as a Surrogate for LF/HF: A Revised Investigation. Modelling and Simulation in Engineering. 2012:8.

Waran N, Randle H. 2017. What we can measure, we can manage: The importance of using robust welfare indicators in Equitation Science. Applied Animal Behaviour Science. 190:74-81.

Warren-Smith AK, McGreevy PD. 2008. Preliminary Investigations Into the Ethological Relevance of Round-Pen (Round-Yard) Training of Horses. Journal of Applied Animal Welfare Science. 11(3):285 - 298.

Warren-Smith AK, Greetham L, McGreevy PD. 2007. Behavioral and physiological responses of horses (Equus caballus) to head lowering. Journal of Veterinary Behavior-Clinical Applications and Research. 2(3):59-67.

Wathan J, Burrows AM, Waller BM, McComb K. 2015. EquiFACS: The Equine Facial Action Coding System. PLOS ONE. 10(8):e0131738.

Webb LE, Veenhoven R, Harfeld JL, Jensen MB. 2018. What is animal happiness? Annals of the New York Academy of Sciences. 1438(1):62-76.

Weiten W. 1992. Psychology: themes and variations. 2nd ed. Pacific Grove, California: Brooks/Cole Publishing Company.

Wells DL. 2009. Sensory stimulation as environmental enrichment for captive animals: A review. Applied Animal Behaviour Science. 118(1-2):1-11.

Wells DL, Hepper PG, Milligan ADS, Barnard S. 2017. Cognitive bias and paw preference in the domestic dog (Canis familiaris). Journal of Comparative Psychology. 131(4):317-325.

Wemelsfelder F. 2007. How animals communicate quality of life: the qualitative assessment of behaviour. Animal Welfare. 16(Supplement 1):25-31.

Wemelsfelder F, Hunter TEA, Mendl MT, Lawrence AB. 2001. Assessing the 'whole animal': a free choice profiling approach. Animal Behaviour. 62:209-220.

Whitaker TC, Goupil X, Roy O, Marciat D, McGahie D. 2011. Evaluation and Comparison Under Field Conditions of the Stress Response Induced in Horses When Administered Endoparasiticides in Tablet or Paste Formulations. International Journal of Applied Research in Veterinary Medicine. 9(1):6-14.

Whittaker AL, Marsh LE. 2019. The role of behavioural assessment in determining 'positive' affective states in animals. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources. 14.

Willson EK, Stratton RB, Bolwell CF, Stafford KJ. 2017. Comparison of positive reinforcement training in cats: A pilot study. Journal of Veterinary Behavior: Clinical Applications and Research. 21:64-70.

Wiśniewska M, Janczarek I, Wilk I, Wnuk-Pawlak E. 2019. Use of Music Therapy in Aiding the Relaxation of Geriatric Horses. Journal of Equine Veterinary Science. 78:89-93.

Wolter R, Stefanski V, Krueger K. 2018. Parameters for the Analysis of Social Bonds in Horses. Animals. 8(11):191.

Yamamuro T, Senzaki K, Iwamoto S, Nakagawa Y, Hayashi T, Hori M, Sakamoto S, Murakami K, Shiga T, Urayama O. 2010. Neurogenesis in the dentate gyrus of the rat hippocampus enhanced by tickling stimulation with positive emotion. Neuroscience Research. 68(4):285-289.

Yáñez-Pizaña A, Mota-Rojas D, Ramírez-Necoechea R, Castillo-Rivera M, Roldán-Santiago P, Mora-Medina P, González-Lozano M. 2019. Application of infrared thermography to assess the effect of different types of environmental enrichment on the ocular, auricular pavilion and nose area temperatures of weaned piglets. Computers and Electronics in Agriculture. 156:33-42.

Yeates JW. 2011. Is 'a life worth living' a concept worth having? Animal Welfare. 20(3):397-406.

Yeates JW, Main DCJ. 2008. Assessment of positive welfare: A review. Veterinary Journal. 175(3):293-300.

Young T, Creighton E, Smith T, Hosie C. 2012. A novel scale of behavioural indicators of stress for use with domestic horses. Applied Animal Behaviour Science. 140(1-2):33-43.

Zebunke M, Langbein J, Manteuffel G, Puppe B. 2011. Autonomic reactions indicating positive affect during acoustic reward learning in domestic pigs. Animal Behaviour. 81(2):481-489.

Zuur A, Ieno E, Walker N, Saveliev A, Smith G. 2009. Mixed effects models and extensions in ecology for R. New York: Springer.

Appendix

A.1. List of related presentations and publications

A.1.1. Presentations

Stratton R, Waran N, Beausoleil N, Stafford K, Worth G, Munn R, Stewart M. 2010a. Non-invasive assessment of positive emotions in horses using behavioural and physiological indicators [Abstract]. In: Hartmann E, Zetterqvist Blokhuis M, Fransson C, Dalin G, editor(s). Proceedings of the 6th International Equitation Science (ISES) Conference; Jul 31-Aug 2; Uppsala, Sweden. p. 42-43.

Stratton R, Waran N, Beausoleil N, Stafford K, Worth G, Munn R, Stewart M. 2010b. Non-invasive assessment of positive emotions in horses using behavioural and physiological indicators [Abstract]. In: Lidifors L, Blokhuis H, Keeling L, editor(s). Proceedings of the 44th Congress of the International Society for Applied Ethology (ISAE); Aug 4-7; Uppsala, Sweden. p. 153-154.

A.1.2. Publications

Beausoleil NJ, Stratton RB, Guesgen MJ, Sutherland MA, Johnson CB. 2016. Scientific evaluation of animal emotions: Brief history and recent New Zealand contributions. Journal of New Zealand Studies. 22:63-71.

Stewart M, Stratton RB, Beausoleil NJ, Stafford KJ, Worth GM, Waran NK. 2011. Assessment of positive emotions in horses: Implications for welfare and performance. Journal of Veterinary Behavior. 6(5):296.

Stratton R, Cogger N, Beausoleil N, Waran N, Stafford K, Stewart M. 2014. Indicators of Good Welfare in Horses. Wellington, New Zealand: New Zealand Government Ministry for Primary Industries. MPI Technical Paper No: 2014/44

A.2. Copyright permissions

CCC Marketplace™

The Royal Society (U.K.) - License Terms and Conditions

This is a License Agreement between Rachael Stratton ("You") and The Royal Society (U.K.) ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by The Royal Society (U.K.), and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	14-Jun-2021
Order License ID	1125668-1
ISSN	1471-2954
Type of Use	Republish in a thesis/dissertation
Publisher	ROYAL SOCIETY
Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Proceedings of the Royal Society of London. Series B: Biological Sciences. Biological sciences
Author/Editor	Royal Society (Great Britain)
Date	01/01/1997
Language	English
Country	United Kingdom of Great Britain and Northern Ireland
Rightsholder	The Royal Society (U.K.)
Publication Type	e-Journal
URL	https://royalsociety.org/journals/

REQUEST DETAILS

Portion Type	Chart/graph/table/figure
Number of charts / graphs / tables / figures requested	1
Format (select all that apply)	Print, Electronic
Who will republish the content?	Academic institution
Duration of Use	Life of current edition
Lifetime Unit Quantity	Up to 499
Rights Requested	Main product
Distribution	Worldwide
Translation	Original language of publication
Copies for the disabled?	Yes
Minor editing privileges?	No
Incidental promotional use?	No
Currency	USD

NEW WORK DETAILS

Title	Positive emotions in animals using the horse as a model
Instructor name	Kevin Stafford
Institution name	Massey University
Expected presentation date	2021-07-30

ADDITIONAL DETAILS

Order reference number	N/A

The requesting person / organization to appear on the license	Rachael Stratton

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1. Core affect represented in two-dimensional space. Words in italics indicate possible locations of specific reported affective states	**Title of the article/chapter the portion is from**	An integrative and functional framework for the study of animal emotion and mood
		Author of portion(s)	Royal Society (Great Britain)
Editor of portion(s)	N/A	**Issue, if republishing an article from a serial**	N/A
Volume of serial or monograph	N/A	**Publication date of portion**	2010-08-04
Page or page range of portion	2896		

PUBLISHER TERMS AND CONDITIONS

Out of Copyright - any journal requested with a Publication date older than 70 years.

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER

THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

   8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

   8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

   8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

   8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

   8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1

AMERICAN PSYCHOLOGICAL ASSOCIATION LICENSE
TERMS AND CONDITIONS

Jun 13, 2021

This Agreement between Rachael Stratton ("You") and American Psychological Association ("American Psychological Association") consists of your license details and the terms and conditions provided by American Psychological Association and Copyright Clearance Center.

License Number	5087290773006
License date	Jun 13, 2021
Licensed Content Publisher	American Psychological Association
Licensed Content Publication	Psychological Bulletin
Licensed Content Title	Ultrasonic vocalizations as indices of affective states in rats.
Licensed copyright line	Copyright © 2002, American Psychological Association
Licensed Content Author	Knutson, Brian; Burgdorf, Jeffrey; Panksepp, Jaak
Licensed Content Date	Nov 1, 2002
Licensed Content Volume	128
Licensed Content Issue	6
Type of Use	Thesis/Dissertation
Requestor type	Academic institution

Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
Rights for	Main product
Duration of use	life of current edition
Creation of copies for the disabled	yes
With minor editing privileges	no
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
Title	Positive emotion in animals using the horse as a model
Institution name	Massey University
Expected presentation date	Jul 2021
Order reference number	Rach
Portions	Figure 1. A two-dimensional map of affective space. Page 962
Requestor Location	Rachael Stratton 73 Laward Road RD4 Palmerston North, other 4474

New Zealand
Attn: Rachael Stratton

Total 0.00 USD

Terms and Conditions

RightsLink Plus Terms and Conditions
American Psychological Association

The publisher for this copyrighted material is the American Psychological Association ("APA"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by the Copyright Clearance Center, Inc. ("CCC") at the time that you opened your CCC account.

APA hereby grants you a non-exclusive license to use this material, subject to the conditions below.

1. Limited License.
   Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed with 12 months from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to a period of 12 months.

2. Credit Lines and Translation Disclaimer.
   The reproduced material must include one of the following credit lines:

   For English language content: Copyright © [year] by [Copyright Owner]. Reproduced [or Adapted] with permission. [Full bibliographic citation.]

   For translated language content: This material originally appeared in English as [full bibliographic citation]. Copyright © [year] by [Copyright Owner]. Translated and reproduced [or adapted] with permission. [Copyright Owner] is not responsible for the quality or accuracy of this translation. This translation cannot be reproduced or distributed further without prior written permission.

3. Author Permission.
   You must obtain the author's (or, in case of multiple authorship, one author's) permission. A good faith effort must be made to locate the author and obtain the author's permission. Author permission is not required for the reuse of small excepts or 1 to 3 figures or 1 to 3 tables; or for course pack or electronic reserve use.

4. Geographic Rights.
   Licenses may be exercised anywhere in the world.

5. Book Edition Limit.
   If the APA content will be reproduced in a book, then permission is limited to the life

of the current edition.

6. Online Use Limitations.
APA full-text articles or book chapters may only be posted on a secure and restricted web site. The credit line specified in Section 2 shall appear on the first screen on which the APA content appears and shall be amended to include the following additional wording: No further reproduction or distribution is permitted without written permission from the American Psychological Association.

7. Reservation of Rights.
APA reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent on Payment.
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by APA or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and APA reserves the right to take any and all action to protect its copyright in the materials. "$0.00" fee means that APA waived the fee.

9. License Revocation.
APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to CCC for denied permissions.

10. Warranties.
APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original to APA and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

11. Indemnity.
You hereby indemnify and agree to hold harmless APA and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. No Transfer of License.
This license is personal to you, but may be assigned or transferred by you to your

employer or a business associate if you give prompt written notice of the assignment or transfer to the APA. No such assignment or transfer shall relieve you of the obligation to pay the designated license fee on a timely basis (although payment by the identified assignee can fulfill your obligation).

13. No Amendment except in Writing.
    This license may not be amended except in a writing signed by both parties (or, in the case of APA, by CCC on APA's behalf).

14. Objection to Contrary Terms.
    APA hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and APA (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

15. Jurisdiction.
    This license transaction shall be governed by and construed in accordance with the laws of the District of Columbia. You hereby agree to submit to the jurisdiction of the federal courts located in the District of Columbia for purposes of resolving any disputes that may arise in connection with this licensing transaction.

Other Conditions:

v1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

A.3. Chapter 5 Supplementary material

A.3.1. Habituation

Table A 1 Percentage of time during each habituation session that each horse spent in each sector location, with summary statistics across tests for each horse.

Group	ID	Name	Statistic	HAB	Sector			
					1	**2**	**3**	**4**
1	1	Shane		1	30	29	34	7
				2	11	30	27	32
			Mean		20.7	29.6	30.1	19.6
			SE		9.7	0.7	3.5	12.6
	3	Lucy		1	4	47	46	3
				2	0	4	96	0
			Mean		1.8	25.6	70.9	1.7
			SE		1.8	21.6	25.1	1.7
	4	Spice		1	7	23	8	62
				2	0	0	100	0
			Mean		3.5	11.3	54.2	31.1
			SE		3.5	11.3	45.8	31.1
2	5	Semitone		1	40	17	40	2
				2	6	89	5	0
			Mean		23.3	53.3	22.3	1.2
			SE		17.2	36.2	17.7	1.2
	6	Phoenix		1	33	24	38	5
				2	4	13	48	35
			Mean		18.2	18.4	43.1	20.3
			SE		14.7	5.2	4.9	15.0
	7	Hattie		1	31	14	41	14
				2	1	15	78	6
			Mean		15.9	14.6	59.2	10.3
			SE		14.7	0.3	18.4	4.0
	8	Streisand		1	7	29	51	13
				2	6	7	72	15
			Mean		6.6	18.1	61.5	13.9
			SE		0.9	10.8	10.5	1.2
3	9	Frankie		1	13	18	8	61

Group	ID	Name	Statistic	HAB	Sector			
					1	2	3	4
				2	0	1	99	0
			Mean		6.5	9.3	53.8	30.4
			SE		6.5	8.7	45.6	30.4
	10	Minty		1	61	16	6	16
				2	29	7	64	0
			Mean		45.4	11.3	35.2	8.1
			SE		16.1	4.6	28.9	8.1
	11	Chonty		1	18	6	35	41
				2	40	60	0	0
			Mean		28.8	33.2	17.8	20.3
			SE		10.9	26.8	17.5	20.3
	12	Whitney		1	4	39	34	22
				2	0	13	52	35
			Mean		2.1	26.0	43.2	28.7
			SE		2.1	13.4	8.8	6.7
4	14	Pixie Fae		1	0	100	0	0
				2	0	0	100	0
			Mean		0.0	50.0	50.0	0.0
			SE		0.0	50.0	50.0	0.0
	15	Toy		1	0	0	100	0
				2	75	25	0	0
			Mean		37.3	12.7	50.0	0.0
			SE		37.3	12.7	50.0	0.0
	16	Grace		1	0	0	21	79
				2	7	47	47	0
			Mean		3.4	23.3	33.9	39.5
			SE		3.4	23.3	12.9	39.5
5	17	32		1	0	71	11	17
				2	0	27	35	38
			Mean		0.2	48.8	23.3	27.8
			SE		0.2	22.2	11.9	10.4
	18	Zoe		1	17	39	27	18
				2	13	15	41	32
			Mean		14.7	26.9	33.8	24.7

Group	ID	Name	Statistic	HAB	Sector			
					1	2	3	4
			SE		2.1	11.9	6.9	7.1
	19	Daisy		1	9	35	26	31
				2	11	11	51	27
			Mean		9.8	22.9	38.5	28.9
			SE		1.1	11.7	12.9	2.3
	20	Fatty		1	0	19	68	13
				2	0	10	62	29
			Mean		0.0	14.2	64.8	21.0
			SE		0.0	4.6	3.1	7.8

ID = Identification number, SE = Standard error of the mean, HAB = Habituation session

A.3.2. Testing

Table A 2 Percentage of time during each test that each horse spent in each sector location, with summary statistics across tests for each horse.

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Sector			
						1	2	3	4
1	1	Shane	SPGC		1	0	0	100	0
					2	0	0	100	0
					3	0	0	100	0
					4	0	0	100	0
				Mean		0.0	0.0	100.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	0.0	100.0	0.0
	3	Lucy	CSPG		1	0	0	100	0
					2	0	0	100	0
					3	0	0	100	0
					4	0	0	100	0
				Mean		0.0	0.0	100.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	0.0	100.0	0.0
	4	Spice	GSPC		1	0	0	100	0
					2	0	0	100	0
					3	0	0	100	0
					4	0	0	100	0

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Sector			
						1	2	3	4
				Mean		0.0	0.0	100.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	0.0	100.0	0.0
2	5	Semitone	PGCS		1	45	55	0	0
					2	0	100	0	0
					3	0	100	0	0
					4	0	100	0	0
				Mean		11.2	88.8	0.0	0.0
				SE		11.2	11.2	0.0	0.0
				Median		0.0	100.0	0.0	0.0
	6	Phoenix	GPCS		1	0	100	0	0
					2	84	16	0	0
					3	0	100	0	0
					4	0	100	0	0
				Mean		21.0	79.0	0.0	0.0
				SE		21.0	21.0	0.0	0.0
				Median		0.0	100.0	0.0	0.0
	7	Hattie	CSPG		1	0	0	100	0
					2	0	100	0	0
					3	0	100	0	0
					4	0	0	23	77
				Mean		0.0	50.0	30.7	19.3
				SE		0.0	28.9	23.7	19.3
				Median		0.0	50.0	11.4	0.0
	8	Streisand	SGPC		1	0	2	5	93
					2	0	3	98	0
					3	0	3	97	0
					4	0	0	4	96
				Mean		0.0	1.8	51.2	47.0
				SE		0.0	0.6	26.7	27.2
				Median		0.0	2.2	51.4	46.3
3	9	Frankie	CPGS		1	0	100	0	0
					2	0	100	0	0
					3	0	82	18	0

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Sector			
						1	2	3	4
				Mean		0.0	93.9	6.1	0.0
				SE		0.0	6.1	6.1	0.0
				Median		0.0	100.0	0.0	0.0
	10	Minty	PSCG		1	87	7	3	4
					2	2	11	72	15
					3	8	19	65	7
				Mean		32.6	12.2	46.4	8.8
				SE		27.3	3.6	22.2	3.4
				Median		8.3	10.6	65.5	7.3
	11	Chonty	GCSP		1	100	0	0	0
					2	0	0	0	100
					3	0	6	63	31
				Mean		33.9	2.0	20.5	43.6
				SE		33.3	2.0	21.0	29.6
				Median		0.0	0.0	0.0	30.8
	12	Whitney	SGPC		1	0	0	100	0
					2	0	0	100	0
					3	0	0	100	0
				Mean		0.0	0.0	100.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	0.0	100.0	0.0
4	14	Pixie Fae	PCSG		1	0	0	0	100
					2	0	0	0	100
					3	0	0	0	100
					4	100	0	0	0
				Mean		25.0	0.0	0.0	75.0
				SE		25.0	0.0	0.0	25.0
				Median		0.0	0.0	0.0	100.0
	15	Toy	GSPC		1	9	0	28	63
					2	0	100	0	0
					3	95	5	0	0
					4	0	100	0	0
				Mean		26.1	51.2	7.0	15.7
				SE		23.1	28.2	7.0	15.7

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Sector			
						1	2	3	4
				Median		4.6	52.4	0.0	0.0
	16	Grace	CPGS		1	0	100	0	0
					2	0	100	0	0
					3	97	3	0	0
					4	0	100	0	0
				Mean		24.3	75.7	0.0	0.0
				SE		24.3	24.3	0.0	0.0
				Median		0.0	100.0	0.0	0.0
5	17	32	CPGS		1	100	0	0	0
					2	35	11	4	50
					3	0	97	3	0
					4	0	0	11	89
				Mean		33.7	27.0	4.5	34.8
				SE		23.6	23.4	2.2	21.6
				Median		17.3	5.7	3.8	24.8
	18	Zoe	PGSC		1	83	17	0	0
					2	0	0	2	98
					3	82	11	7	0
					4	10	6	75	9
				Mean		43.9	8.3	21.1	26.7
				SE		22.4	3.6	18.0	23.8
				Median		46.1	8.3	4.8	4.4
	19	Daisy	GSCP		1	0	4	96	0
					2	0	20	80	0
					3	0	0	81	19
					4	0	0	100	0
				Mean		0.0	5.8	89.4	4.8
				SE		0.0	4.7	5.2	4.8
				Median		0.0	1.8	88.6	0.0
	20	Fatty	SCPG		1	0	100	0	0
					2	0	0	100	0
					3	0	0	4	96
					4	0	0	3	97
				Mean		0.0	25.0	26.7	48.3

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Sector			
						1	2	3	4
				SE		0.0	25.0	24.4	27.9
				Median		0.0	0.0	3.5	48.1

ID = Identification number, SE = Standard error of the mean, G = Groom, P = Person, S = Spray, C = No-stimulus

[a] The randomly allocated stimulus location across sectors 1-4. For example, for horse ID 1, Shane, Spray was associated with sector 1, Person with sector 2, Groom with sector 3, and No-stimulus with sector 4.

Table A 3 Percentage of time during each test that each horse spent in each stimulus location, with summary statistics across tests for each horse.

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Stimulus			
						Groom	Person	Spray	No-stimulus
1	1	Shane	SPGC		1	100	0	0	0
					2	100	0	0	0
					3	100	0	0	0
					4	100	0	0	0
				Mean		100.0	0.0	0.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		100.0	0.0	0.0	0.0
	3	Lucy	CSPG		1	0	100	0	0
					2	0	100	0	0
					3	0	100	0	0
					4	0	100	0	0
				Mean		0.0	100.0	0.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	100.0	0.0	0.0
	4	Spice	GSPC		1	0	100	0	0
					2	0	100	0	0
					3	0	100	0	0
					4	0	100	0	0
				Mean		0.0	100.0	0.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	100.0	0.0	0.0
2	5	Semitone	PGCS		1	55	45	0	0
					2	100	0	0	0
					3	100	0	0	0
					4	100	0	0	0
				Mean		88.8	11.2	0.0	0.0
				SE		11.2	11.2	0.0	0.0
				Median		100.0	0.0	0.0	0.0

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Stimulus			
						Groom	Person	Spray	No-stimulus
	6	Phoenix	GPCS		1	0	100	0	0
					2	84	16	0	0
					3	0	100	0	0
					4	0	100	0	0
				Mean		21.0	79.0	0.0	0.0
				SE		21.0	21.0	0.0	0.0
				Median		0.0	100.0	0.0	0.0
	7	Hattie	CSPG		1	0	100	0	0
					2	0	0	100	0
					3	0	0	100	0
					4	77	23	0	0
				Mean		19.3	30.7	50.0	0.0
				SE		19.3	23.7	28.9	0.0
				Median		0.0	11.4	50.0	0.0
	8	Streisand	SGPC		1	2	5	0	93
					2	3	98	0	0
					3	3	97	0	0
					4	0	4	0	96
				Mean		1.8	51.2	0.0	47.0
				SE		0.6	26.7	0.0	27.2
				Median		2.2	51.4	0.0	46.3
3	9	Frankie	CPGS		1	0	100	0	0
					2	0	100	0	0
					3	18	82	0	0
				Mean		6.1	93.9	0.0	0.0
				SE		6.1	6.1	0.0	0.0
				Median		0.0	100.0	0.0	0.0
	10	Minty	PSCG		1	4	87	7	3
					2	15	2	11	72
					3	7	8	19	65
				Mean		9.0	32.2	12.1	46.8
				SE		3.4	27.3	3.6	22.2
				Median		7.3	8.3	10.6	65.5

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Stimulus			
						Groom	Person	Spray	No-stimulus
	11	Chonty	GCSP		1	100	0	0	0
					2	0	100	0	0
					3	0	31	63	6
				Mean		33.3	43.6	21.0	2.0
				SE		33.3	29.6	21.0	2.0
				Median		0.0	30.8	0.0	0.0
	12	Whitney	SGPC		1	0	100	0	0
					2	0	100	0	0
					3	0	100	0	0
				Mean		0.0	100.0	0.0	0.0
				SE		0.0	0.0	0.0	0.0
				Median		0.0	100.0	0.0	0.0
4	14	Pixie Fae	PCSG		1	100	0	0	0
					2	100	0	0	0
					3	100	0	0	0
					4	0	100	0	0
				Mean		75.0	25.0	0.0	0.0
				SE		25.0	25.0	0.0	0.0
				Median		100.0	0.0	0.0	0.0
	15	Toy	GSPC		1	9	28	0	63
					2	0	0	100	0
					3	95	0	5	0
					4	0	0	100	0
				Mean		26.1	7.0	51.2	15.7
				SE		23.1	7.0	28.2	15.7
				Median		4.6	0.0	52.4	0.0
	16	Grace	CPGS		1	0	100	0	0
					2	0	100	0	0
					3	0	3	0	97
					4	0	100	0	0
				Mean		0.0	75.7	0.0	24.3
				SE		0.0	24.3	0.0	24.3
				Median		0.0	100.0	0.0	0.0

Group	ID	Name	Stimulus sector allocation [a]	Statistic	Test	Stimulus			
						Groom	Person	Spray	No-stimulus
5	17	32	CPGS		1	0	0	0	100
					2	4	11	50	35
					3	3	97	0	0
					4	11	0	89	0
				Mean		4.5	27.0	34.8	33.7
				SE		2.2	23.4	21.6	23.6
				Median		3.8	5.7	24.8	17.3
	18	Zoe	PGSC		1	17	83	0	0
					2	0	0	2	98
					3	11	82	7	0
					4	6	10	75	9
				Mean		8.3	43.9	21.1	26.7
				SE		3.6	22.4	18.0	23.8
				Median		8.3	46.1	4.8	4.4
	19	Daisy	GSCP		1	0	0	4	96
					2	0	0	20	80
					3	0	19	0	81
					4	0	0	0	100
				Mean		0.0	4.8	5.8	89.4
				SE		0.0	4.8	4.7	5.2
				Median		0.0	0.0	1.8	88.6
	20	Fatty	SCPG		1	0	0	0	100
					2	0	100	0	0
					3	96	4	0	0
					4	97	3	0	0
				Mean		48.3	26.7	0.0	25.0
				SE		27.9	24.4	0.0	25.0
				Median		48.1	3.5	0.0	0.0

ID = Identification number, SE = Standard error of the mean, G = Groom, P = Person, S = Spray, C = No-stimulus

[a] The randomly allocated stimulus location across sectors 1-4. For example, for horse ID 1, Shane, Spray was associated with sector 1, Person with sector 2, Groom with sector 3, and No-stimulus with sector 4.

A.4. Chapter 7 Supplementary material

A.4.1. Heart rate and heart rate variability

Heart rate

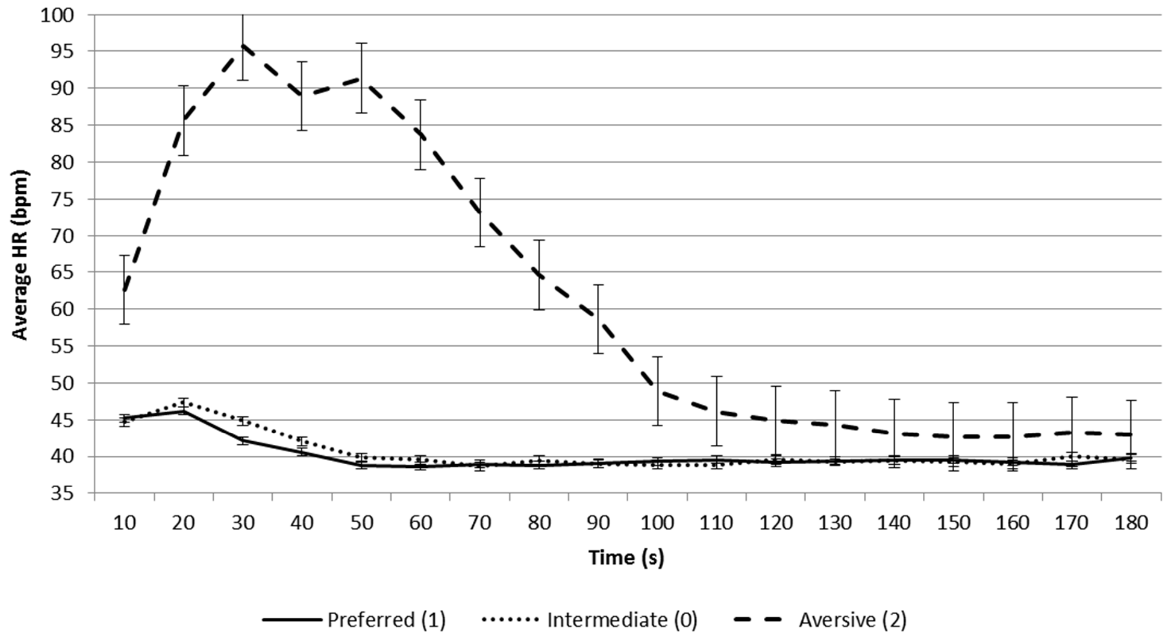


Figure A 1 Observed heart rate, averaged every 10 seconds, over time from the start of treatment stratified by preference for a treatment.
Bars show standard error (SE). bpm = beats per minute, s = seconds.

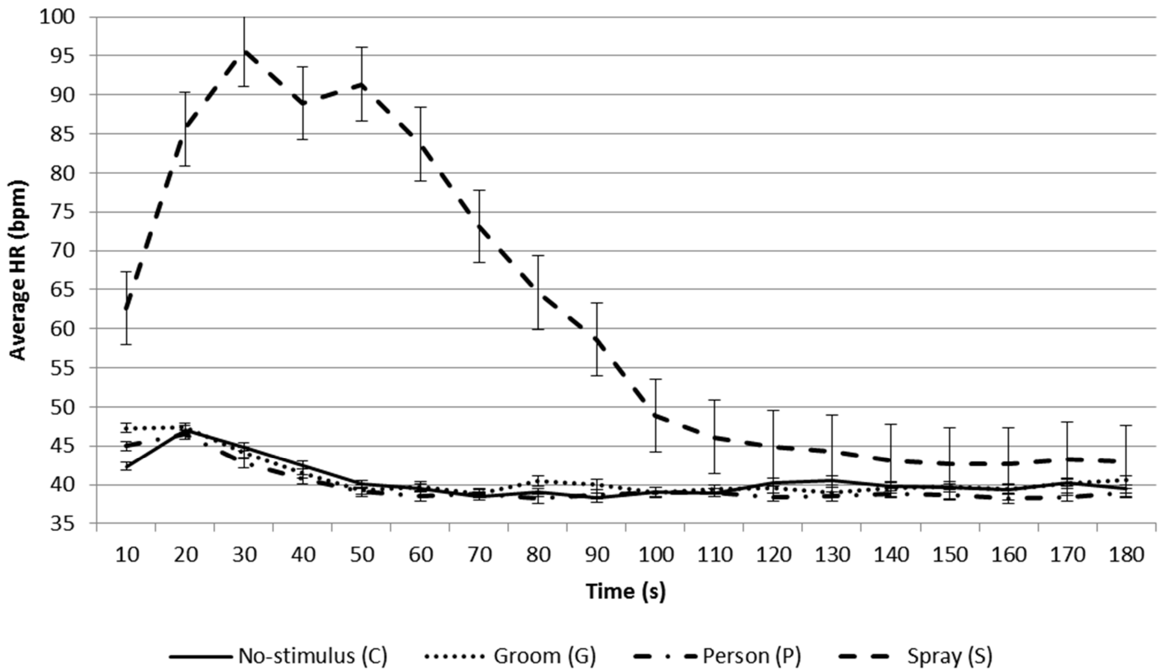


Figure A 2 Observed heart rate, averaged every 10 seconds, over time from the start of treatment stratified by treatment.
Bars show standard error (SE). bpm = beats per minute, s = seconds.

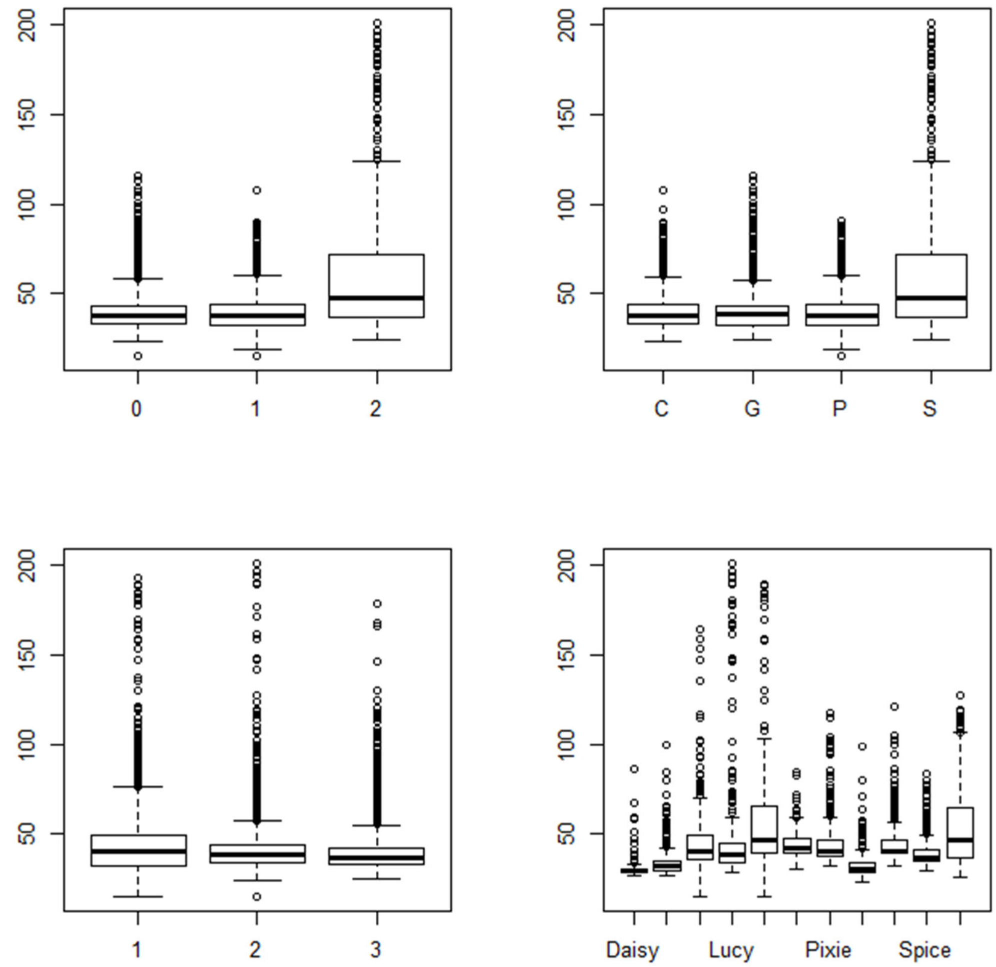


Figure A 3 Boxplots showing the heart rate data in beats per minute stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Inter-beat interval

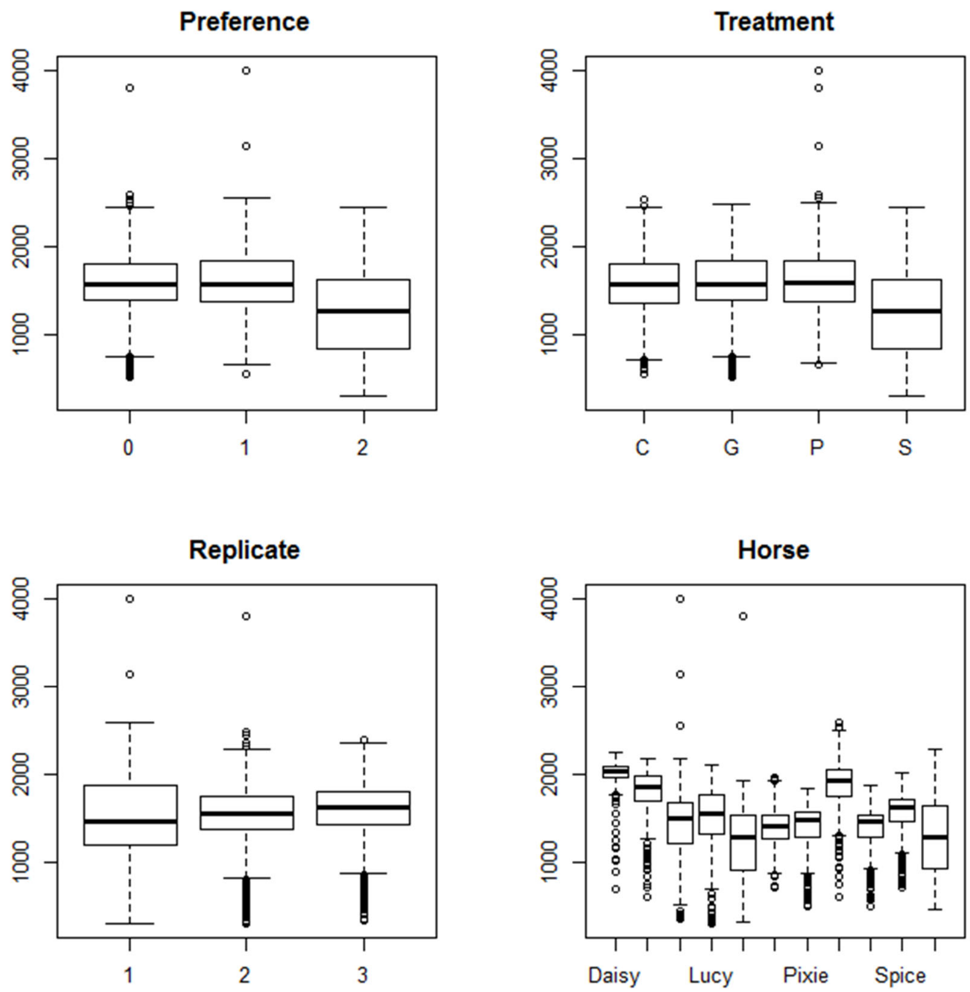


Figure A 4 Boxplots showing the inter-beat interval data in milliseconds stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

SDNN

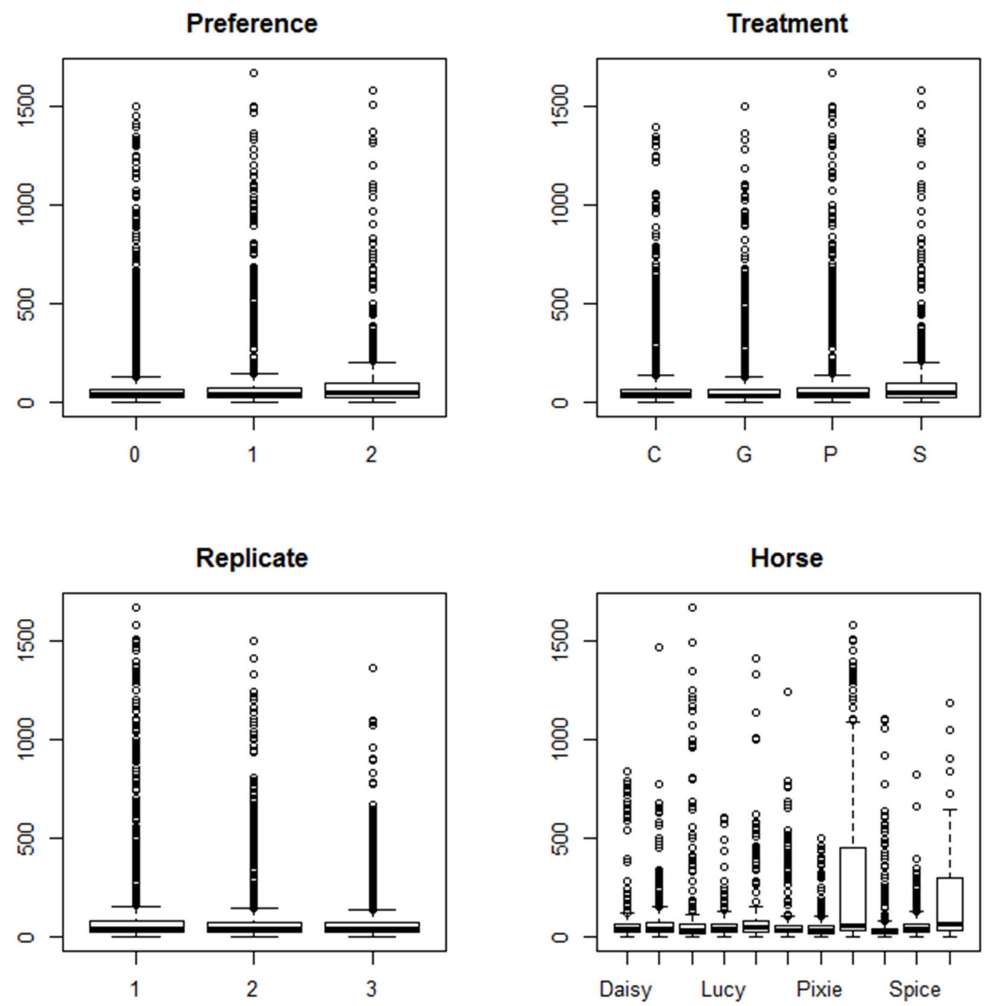


Figure A 5 Boxplots showing the SDNN data in milliseconds stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

RMSSD

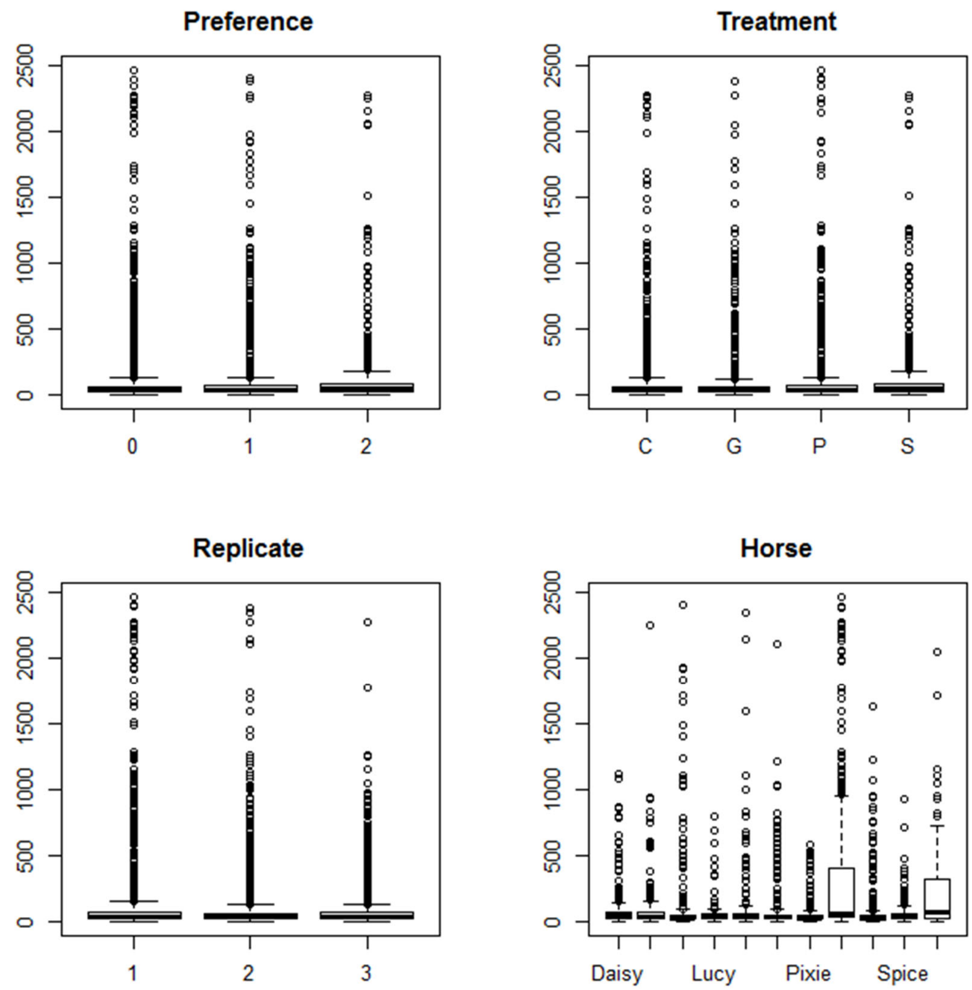


Figure A 6 Boxplots showing the RMSSD data in milliseconds stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

RMSSD:SDNN ratio

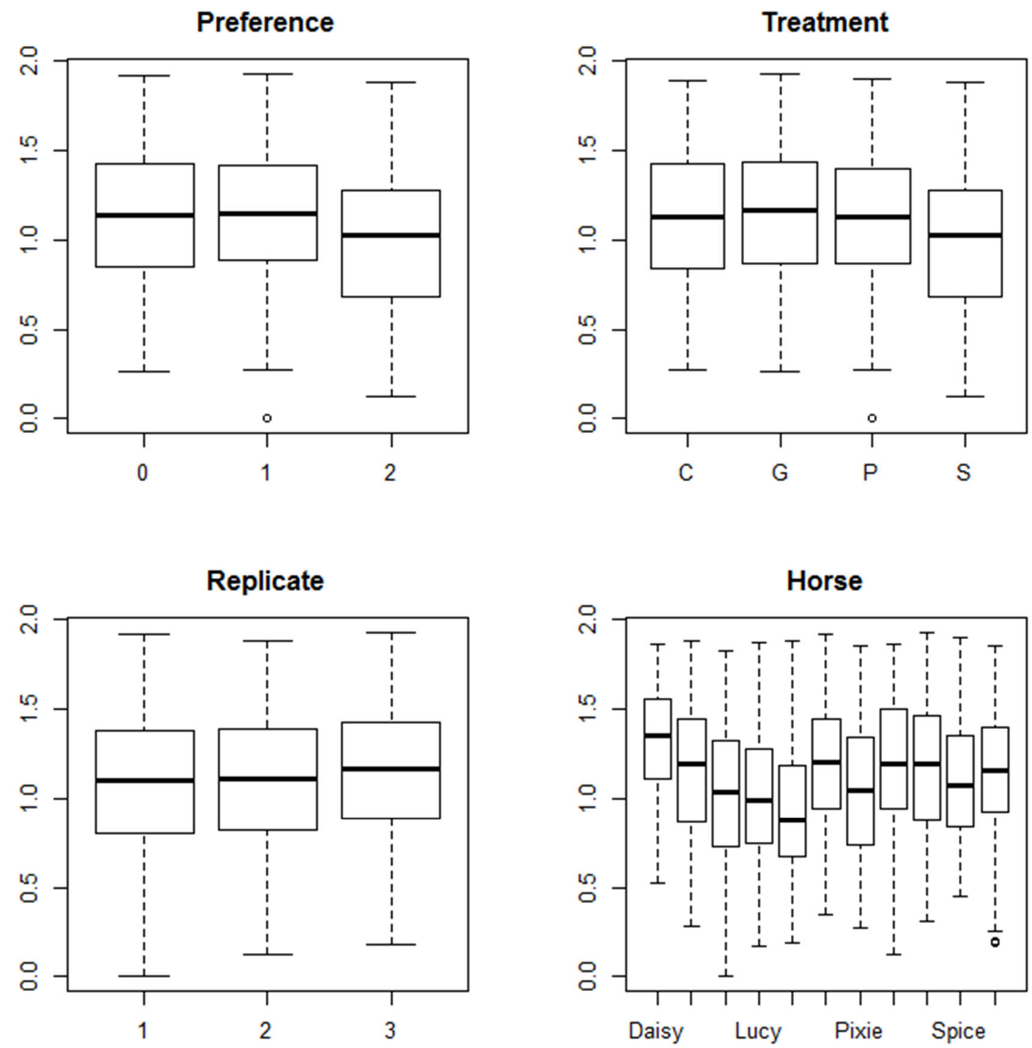


Figure A 7 Boxplots showing the RMSSD:SDNN ratio data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Analysis by preference

Table A 4 Results from generalised least square regression models describing the relationship between five parameters for heart rate variability and preference for a treatment.
The intercept equates to the value during the Preferred treatment as there was only one variable in the model. Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean
LogHRav (bpm)	Intercept (Preferred)	3.68	0.05	<0.001	39.65
	Intermediate	0.01	0.001	0.593	39.92
	Aversive	***0.23[a]***	***0.02***	***<0.001****	***49.90***
IBIav (ms)	Intercept (Preferred)	1575.88	76.58	<0.001	1575.88
	Intermediate	-14.39	15.13	0.342	1561.49
	Aversive	***-284.30***	***21.28***	***<0.001****	***1291.58***
LogSDDN (ms)	Intercept (Preferred)	3.97	0.10	<0.001	51.98
	Intermediate	-0.06	0.04	0.152	48.9
	Aversive	**0.14**	**0.06**	**0.019***	**59.95**
LogRMSSD (ms)	Intercept (Preferred)	4.05	0.11	<0.001	56.40
	Intermediate	-0.07	0.04	0.084	52.52
	Aversive	-0.04	0.06	0.509	54.15
Ratio RMSSD/SDNN	Intercept (Preferred)	1.14	0.03	<0.001	1.14
	Intermediate	-0.01	0.01	0.486	1.13
	Aversive	***-0.15***	***0.02***	***<0.001****	***0.99***

* p ≤ 0.05
[a] The logHRav for the aversive was 0.23 higher than during the preferred treatment. That is the predicted logHRav is 3.91 (3.68 + 0.23), which was statistically significant (p < 0.001).

Analysis by treatment

Table A 5 Results from generalised least square regression models describing the relationship between five parameters for heart rate variability and treatment. The intercept equates to the value during the No-stimulus treatment as there was only one variable in the model. Analysis is based on comparison to the No-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean
LogHrav (bpm)	Intercept (No-stimulus)	3.66	0.05	<0.001	38.86
	Groom	***0.05***[a]	***0.01***	***<0.001****	***40.85***
	Person	0.00	0.01	0.997	38.86
	Spray	***0.25***	***0.02***	***<0.001****	***49.90***
IBIav (ms)	Intercept (No-stimulus)	1581.08	76.58	<0.0001	1581.08
	Groom	***-52.84***	***17.60***	***0.003****	***1528.24***
	Person	8.38	17.54	0.633	1589.46
	Spray	***-289.96***	***21.67***	***<0.001****	***1291.12***
LogSDDN (ms)	Intercept (No-stimulus)	3.91	0.1	<0.001	48.90
	Groom	-0.00	0.05	0.973	48.80
	Person	0.07	0.05	0.156	52.52
	Spray	***0.20***	***0.06***	***<0.001****	***60.07***
LogRMSSD (ms)	Intercept (No-stimulus)	3.97	0.11	<0.001	51.98
	Groom	0.01	0.05	0.780	52.52
	Person	0.07	0.05	0.146	55.83
	Spray	0.04	0.06	0.536	54.15
Ratio RMSSD/SDNN	Intercept (No-stimulus)	1.13	0.03	<0.001	1.13
	Groom	0.02	0.02	0.223	1.15
	Person	0.00	0.02	0.950	1.13
	Spray	***-0.14***	***0.02***	***<0.001****	***0.99***

* p ≤ 0.05
[a] The logHRav was 0.05 significantly higher while being Groomed than during the No-stimulus treatment (p < 0.001). The model predicts that logHRav for groom treatment is 3.71 (3.66 + 0.05).

A.4.2. Eye Temperature

Average eye temperature

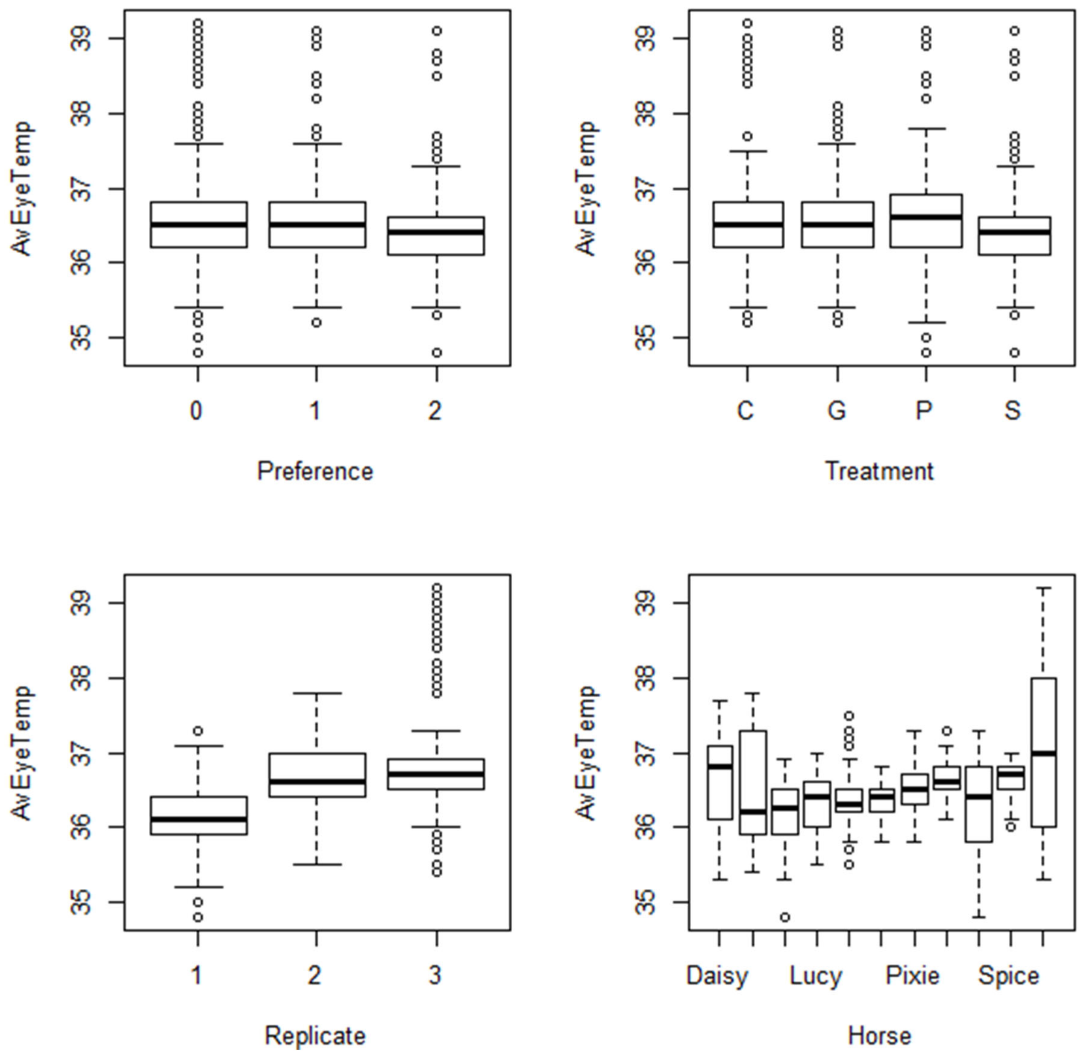


Figure A 8 Boxplots showing the average eye temperature (AvEyeTemp) data in ºC stratified by (clockwise from top left) preference, treatment, horse, and replicate. The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Maximum eye temperature

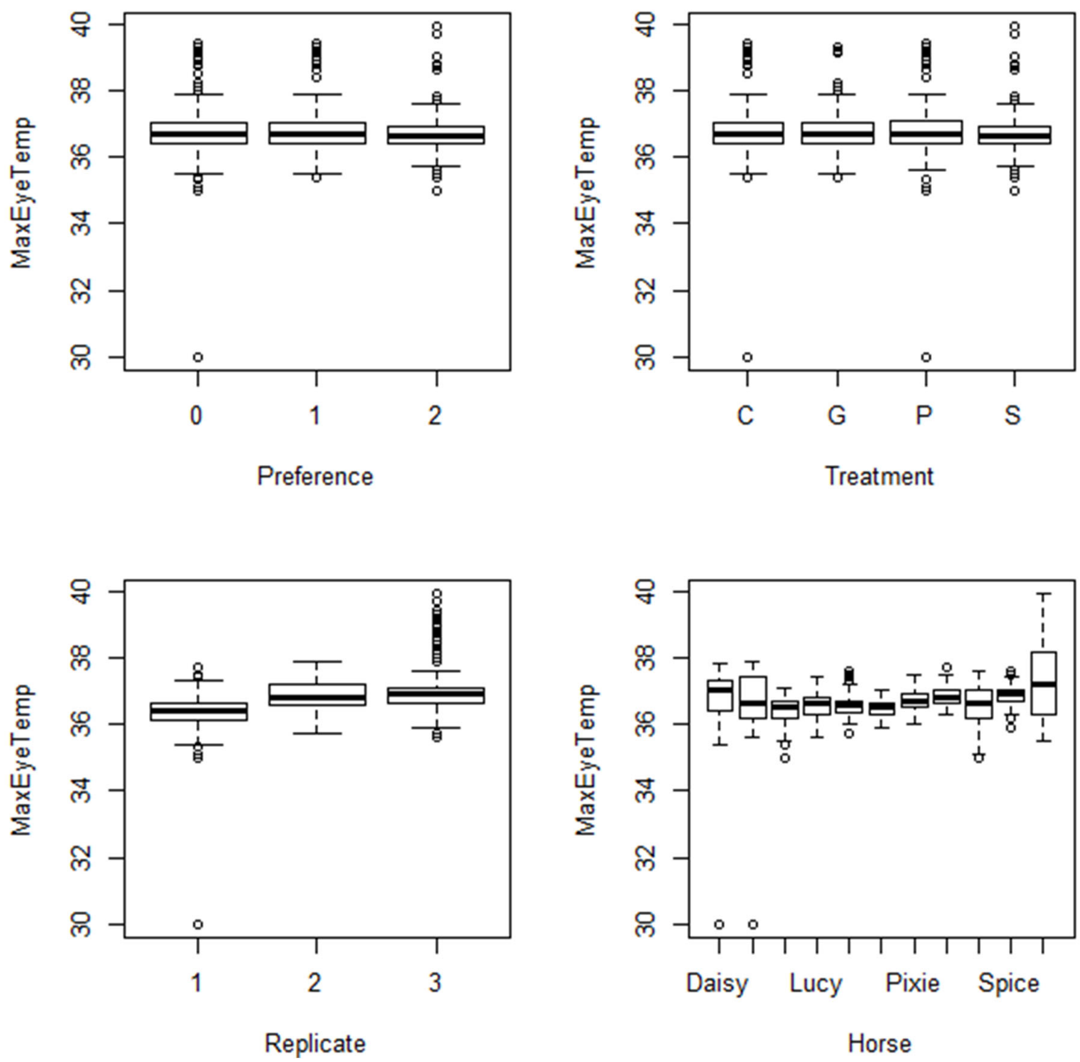


Figure A 9 Boxplots showing the maximum eye temperature (MaxEyeTemp) data in ºC stratified by (clockwise from top left) preference, treatment, horse, and replicate. The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Analysis by preference

Table A 6 Results of a mixed effects regression model examining the effect of preference for a treatment on average and maximum eye temperature (°C). Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean^
Average eye temperature	Intercept (Preferred)	36.05 [b]	0.17	<0.001	36.05
	Intermediate	-0.01	0.01	0.635	36.04
	Aversive	***-0.07*** [c]	***0.02***	***0.001****	***35.98***
	Initial ambient air temperature [a]	***0.08*** [d]	***0.02***	***0.004****	**-**
Maximum eye temperature	Intercept (Preferred)	36.29	0.17	<0.001	36.29
	Intermediate	-0.01	0.02	0.587	36.28
	Aversive	-0.03	0.03	0.331	36.26
	Initial ambient air temperature [a]	***0.07***	***0.02***	***0.007****	**-**

^ Model predicted mean at 15°C air temperature
* p ≤ 0.05
[a] Ambient air temperature centred at 15°C.
[b] When the preferred treatment was applied and the ambient air temperature was 15°C, the predicted average eye temperature was 36.05°C.
[c] The average eye temperature during the aversive treatment was 0.07°C less than during the preferred treatment, that is the predicted average eye temperature at 15°C was 35.08°C (36.05-0.07). The difference was statistically significant (p = 0.001).
[d] The average eye temperature increased by 0.08°C with every degree increase in initial ambient air temperature. This effect may not be biologically relevant as ambient air temperature was only recorded at the start of each session.

Analysis by treatment

Table A 7 Results of a mixed effects regression model examining the effect of treatment on average eye temperature (°C).
Analysis is based on comparison to the no-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Parameter	Variable	Beta	SE	p-value (Wald)	Predicted mean^
Average eye temperature	Intercept (No-stimulus)	36.04 [b]	0.17	<0.001	36.04
	Groom	-0.02	0.02	0.162	36.02
	Person	**0.04**	**0.02**	**0.025***	**36.08**
	Spray	***-0.06*** [c]	***0.02***	***0.003****	***35.98***
	Initial ambient air temperature [a]	***0.08*** [d]	***0.02***	***0.004****	-
Maximum eye temperature	Intercept (No-stimulus)	36.28	0.17	<0.001	36.28
	Groom	-0.01	0.02	0.726	36.27
	Person	0.02	0.02	0.409	36.30
	Spray	-0.02	0.03	0.569	36.26
	Initial ambient air temperature [a]	***0.07***	***0.02***	***0.007****	-

^ Model predicted mean at 15°C air temperature
* p ≤ 0.05
[a] Ambient air temperature centred at 15°C.
[b] When the No-stimulus treatment was applied and ambient air temperature was 15°C, the average eye temperature was 36.04°C.
[c] The average eye temperature during the Spray treatment was 0.06°C less than during the No-stimulus treatment, that is the predicted average eye temperature during the Spray treatment at 15°C was 35.08°C (36.04-0.06). The difference was statistically significant (p = 0.003).
[d] The average eye temperature increased by 0.08°C with every degree increase in initial ambient air temperature. This effect may not be biologically relevant as ambient air temperature was only recorded at the start of each session.

A.5. Chapter 8 Supplementary material

A.5.1. Foreleg

Foreleg lift

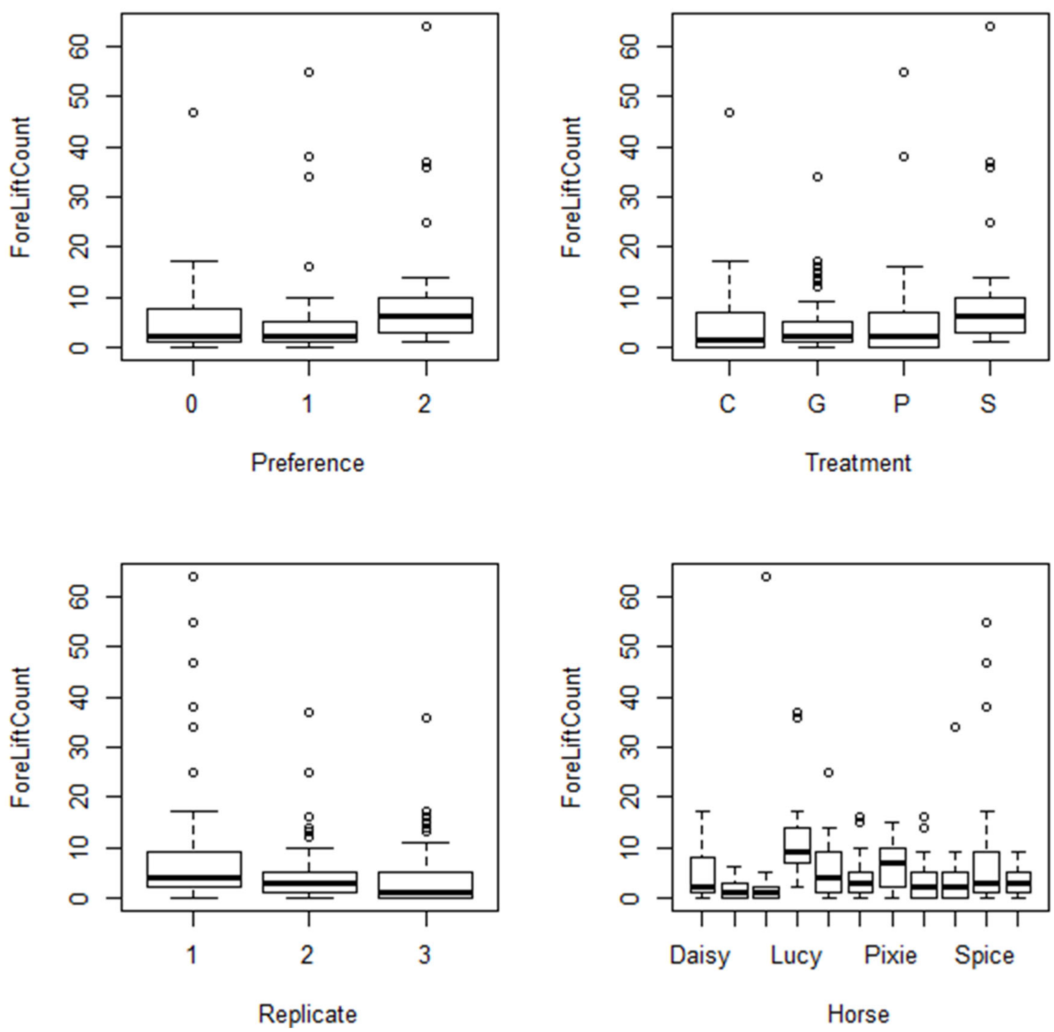


Figure A 10 Boxplots showing the foreleg lift count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Strike

Table A 8 Two-way table showing the number of times a foreleg strike did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Strike (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	63	3	66	5
	0	123	9	132	7
	2	30	3	33	9
Treatment	C	60	6	66	9
	G	64	2	66	3
	P	62	4	66	6
	S	30	3	33	9
Replicate	1	67	10	77	13
	2	72	5	77	6
	3	77	0	77	0
Horse	Daisy	20	1	21	5
	Frankie	21	0	21	0
	Grace	19	2	21	10
	Lucy	20	1	21	5
	Minty	21	0	21	0
	Phoenix	18	3	21	14
	Pixie Fae	20	1	21	5
	Semitone	18	3	21	14
	Shane	21	0	21	0
	Spice	17	4	21	19
	Whitney	21	0	21	0
Total		**216**	**15**	**231**	**6**

Rear

Table A 9 Two-way table showing the number of times a rear did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Rear (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	132	0	132	0
	2	28	5	33	15
Treatment	C	66	0	66	0
	G	66	0	66	0
	P	64	2	66	3
	S	28	5	33	15
Replicate	1	74	3	77	4
	2	75	2	77	3
	3	75	2	77	3
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	20	1	21	5
	Lucy	20	1	21	5
	Minty	20	1	21	5
	Phoenix	21	0	21	0
	Pixie Fae	20	1	21	5
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	19	2	21	10
	Whitney	20	1	21	5
Total		**224**	**7**	**231**	**3**

A.5.2. Hindleg

Hindleg lift

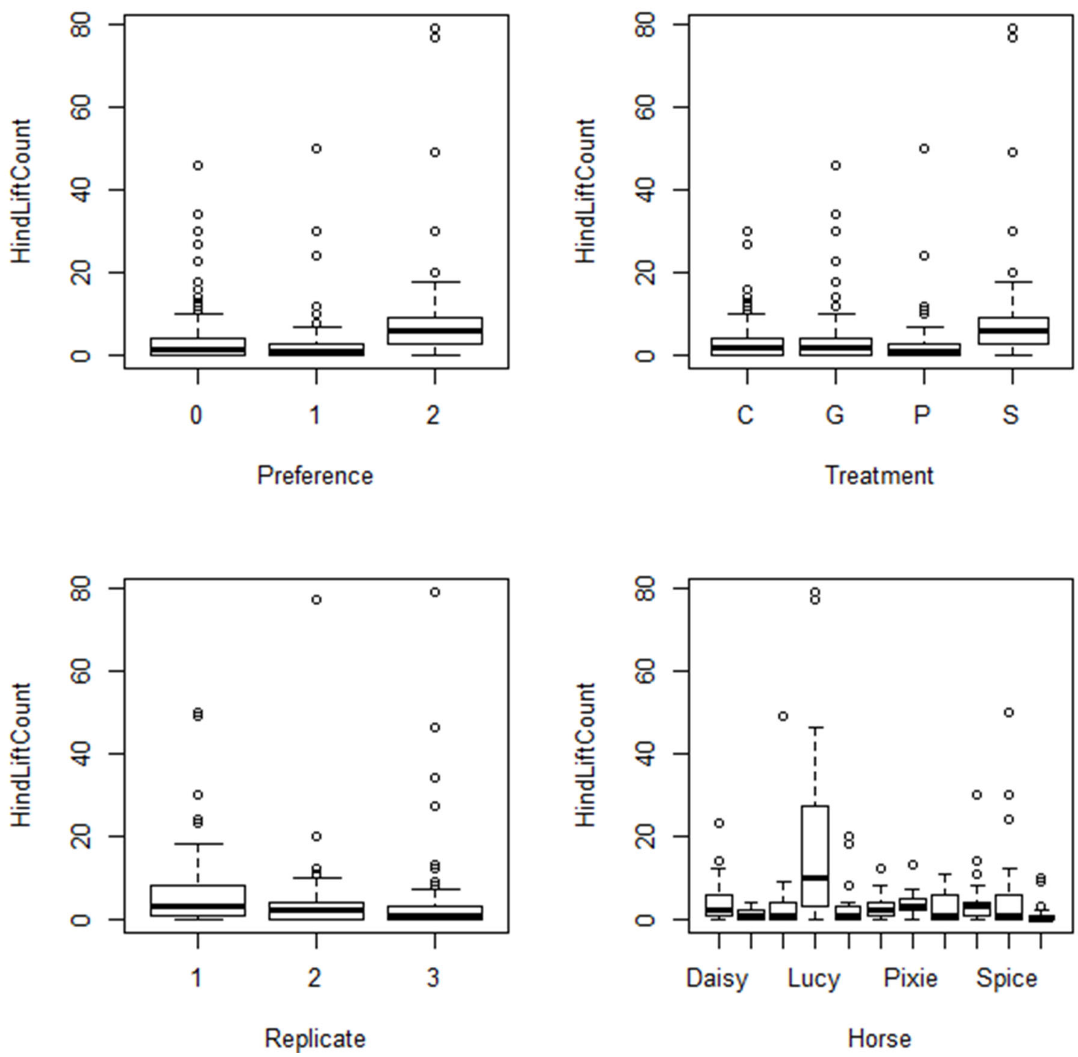


Figure A 11 Boxplots showing the hindleg lift count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Kick

Table A 10 Two-way table showing the number of times a hindleg kick did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Kick (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	125	7	132	5
	2	27	6	33	18
Treatment	C	65	1	66	2
	G	60	6	66	9
	P	64	2	66	3
	S	27	6	33	18
Replicate	1	68	9	77	12
	2	77	0	77	0
	3	71	6	77	8
Horse	Daisy	19	2	21	10
	Frankie	21	0	21	0
	Grace	19	2	21	10
	Lucy	17	4	21	19
	Minty	20	1	21	5
	Phoenix	21	0	21	0
	Pixie Fae	18	3	21	14
	Semitone	21	0	21	0
	Shane	19	2	21	10
	Spice	21	0	21	0
	Whitney	20	1	21	5
Total		**216**	**15**	**231**	**6**

Rest

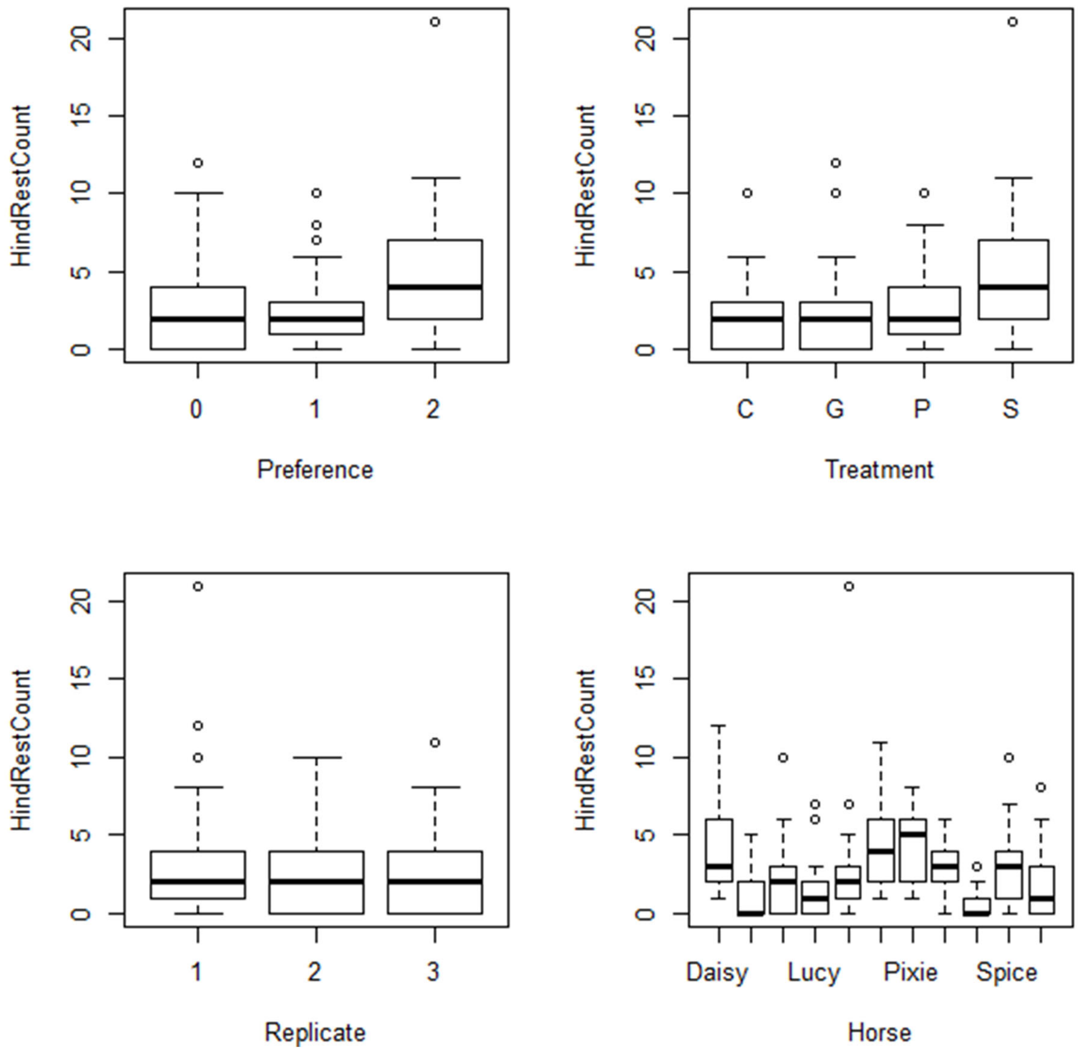


Figure A 12 Boxplots showing the hindleg rest count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Stand

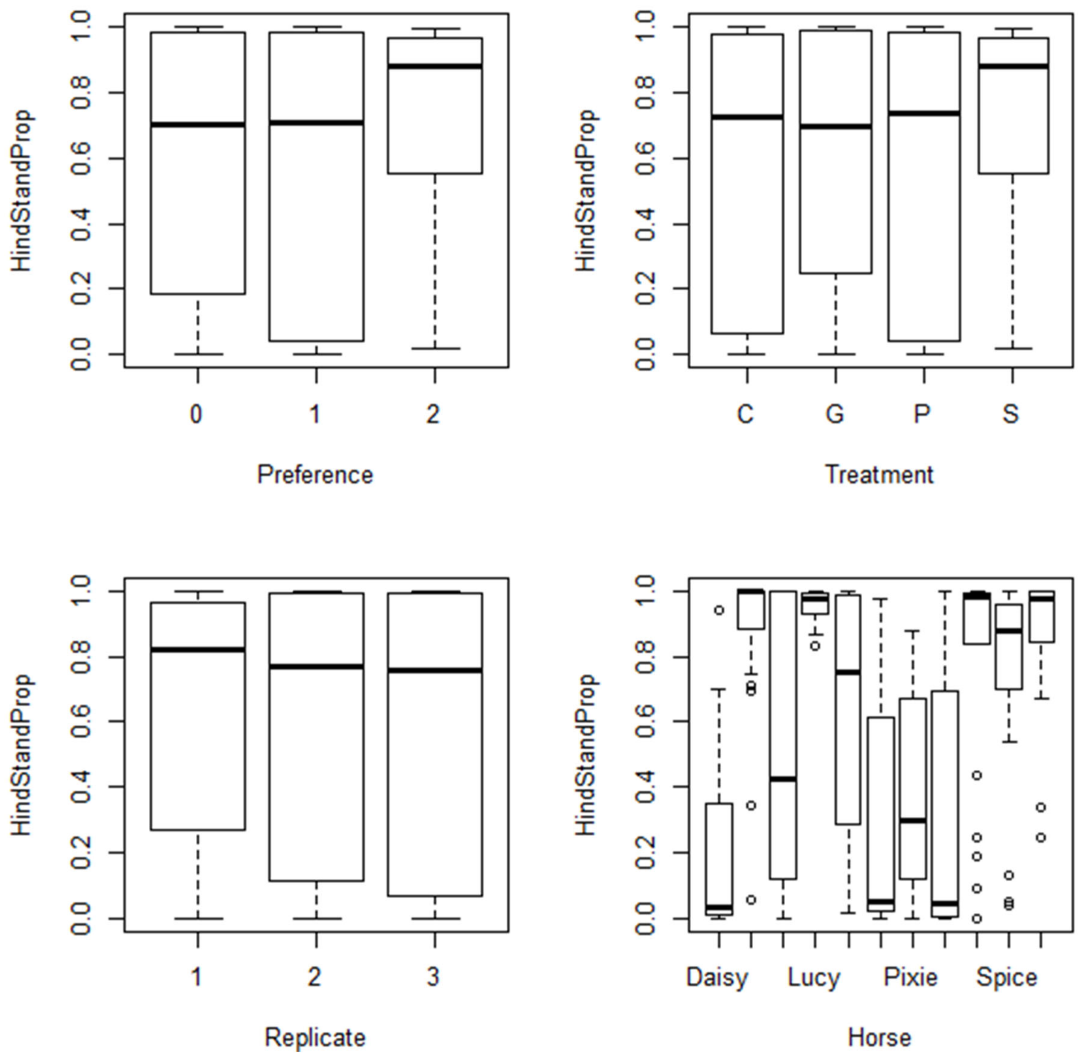


Figure A 13 Boxplots showing the hindleg stand proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

A.5.3. Neck position

Very low

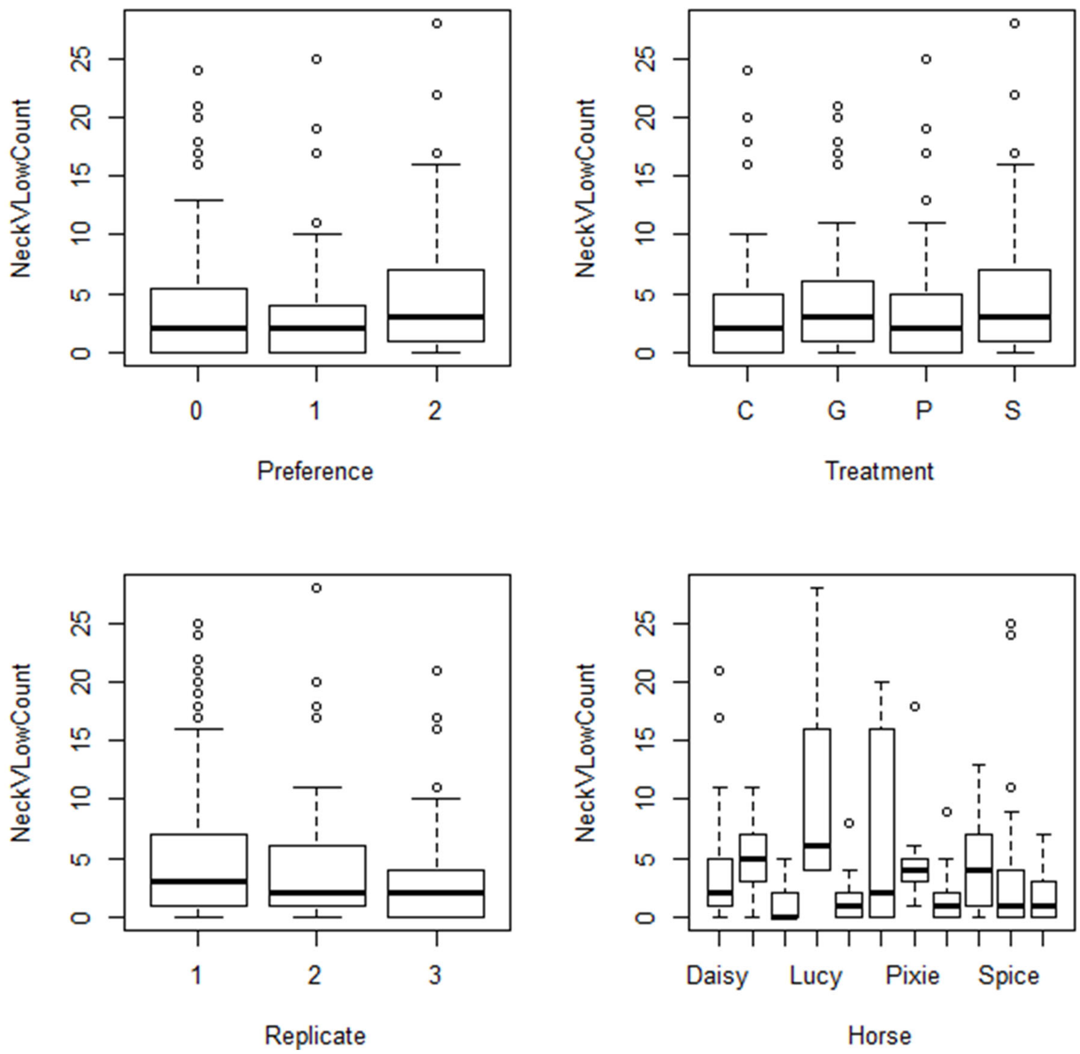


Figure A 14 Boxplots showing the neck very low count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Low

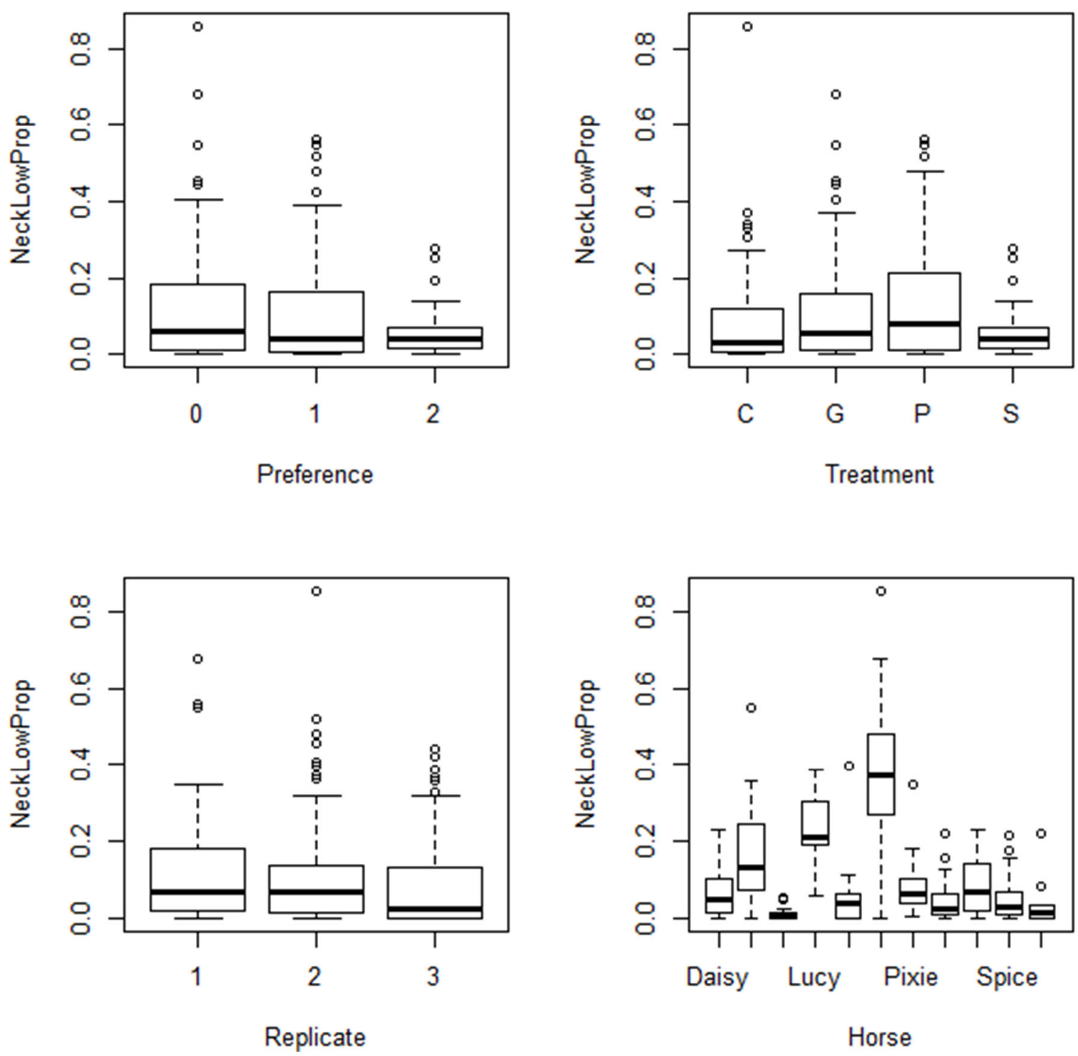


Figure A 15 Boxplots showing the neck low proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Neck Neutral

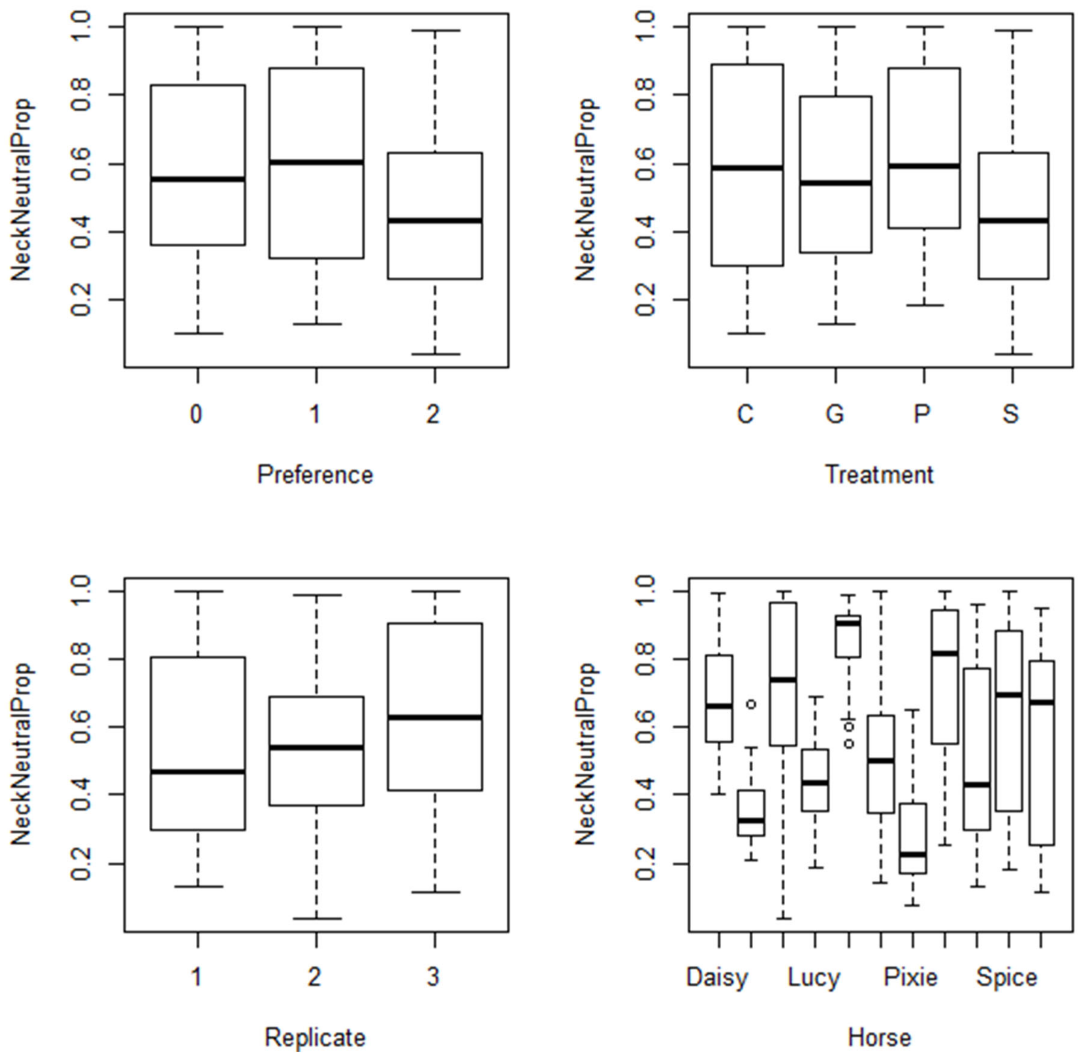


Figure A 16 Boxplots showing the neck neutral proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

High

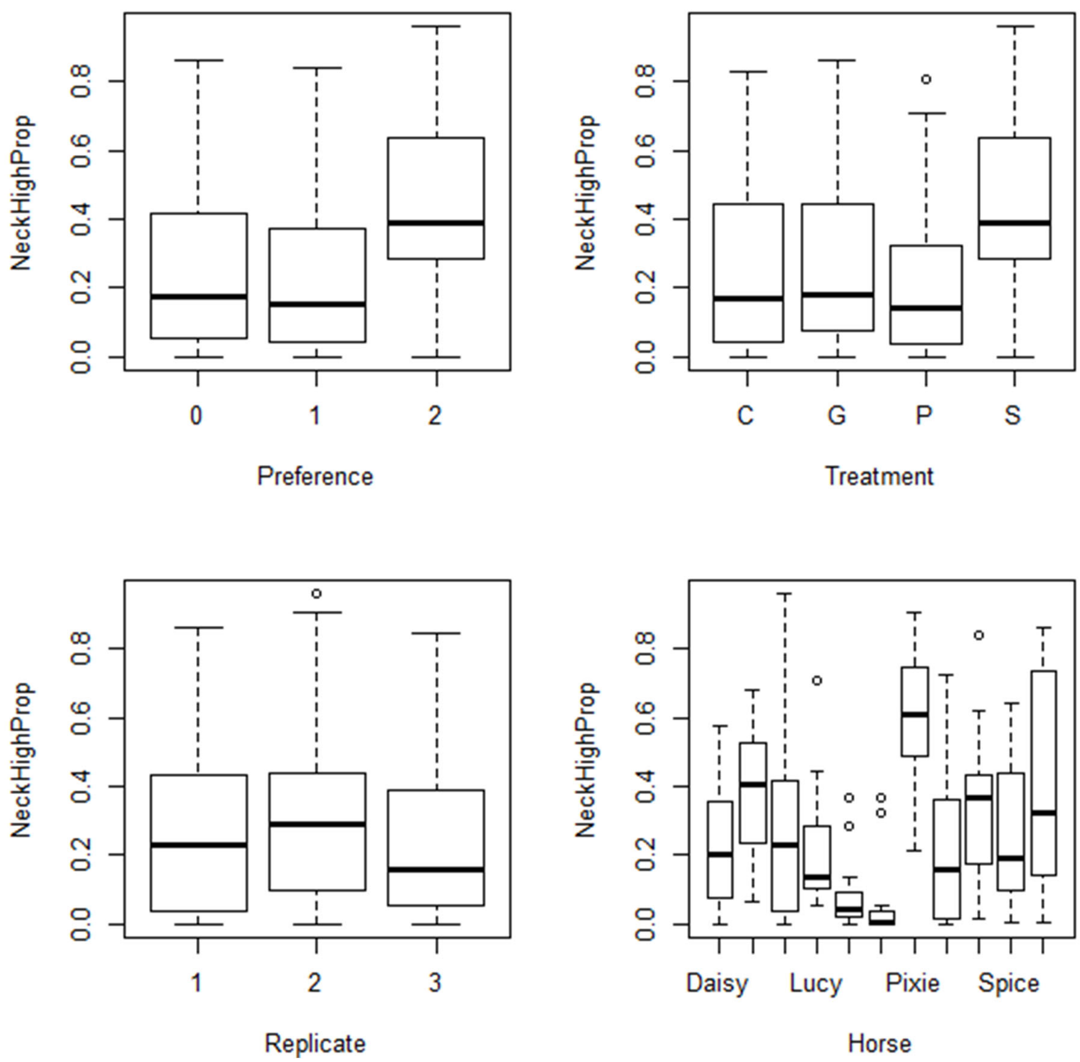


Figure A 17 Boxplots showing the neck high proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Very high

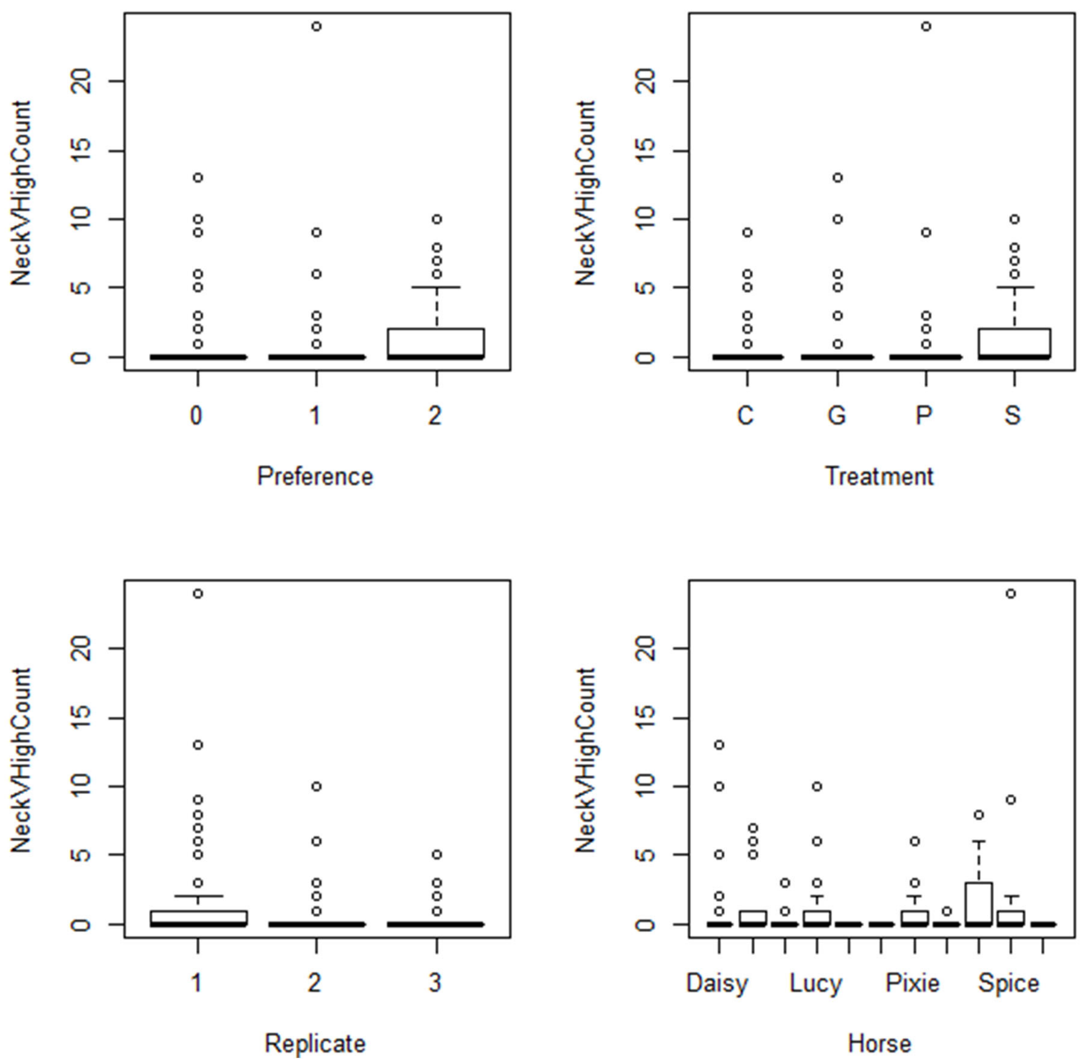


Figure A 18 Boxplots showing the neck very count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray

A.5.4. Tail

Lift

Table A 11 Two-way table showing the number of times a tail lift did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Tail Lift (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	48	18	66	27
	0	92	40	132	30
	2	23	10	33	30
Treatment	C	42	24	66	36
	G	46	20	66	30
	P	52	14	66	21
	S	23	10	33	30
Replicate	1	54	23	77	30
	2	59	18	77	23
	3	50	27	77	35
Horse	Daisy	13	8	21	38
	Frankie	19	2	21	10
	Grace	14	7	21	33
	Lucy	7	14	21	67
	Minty	14	7	21	33
	Phoenix	17	4	21	19
	Pixie Fae	19	2	21	10
	Semitone	14	7	21	33
	Shane	11	10	21	48
	Spice	19	2	21	10
	Whitney	16	5	21	24
Total		**163**	**68**	**231**	**29**

Swish

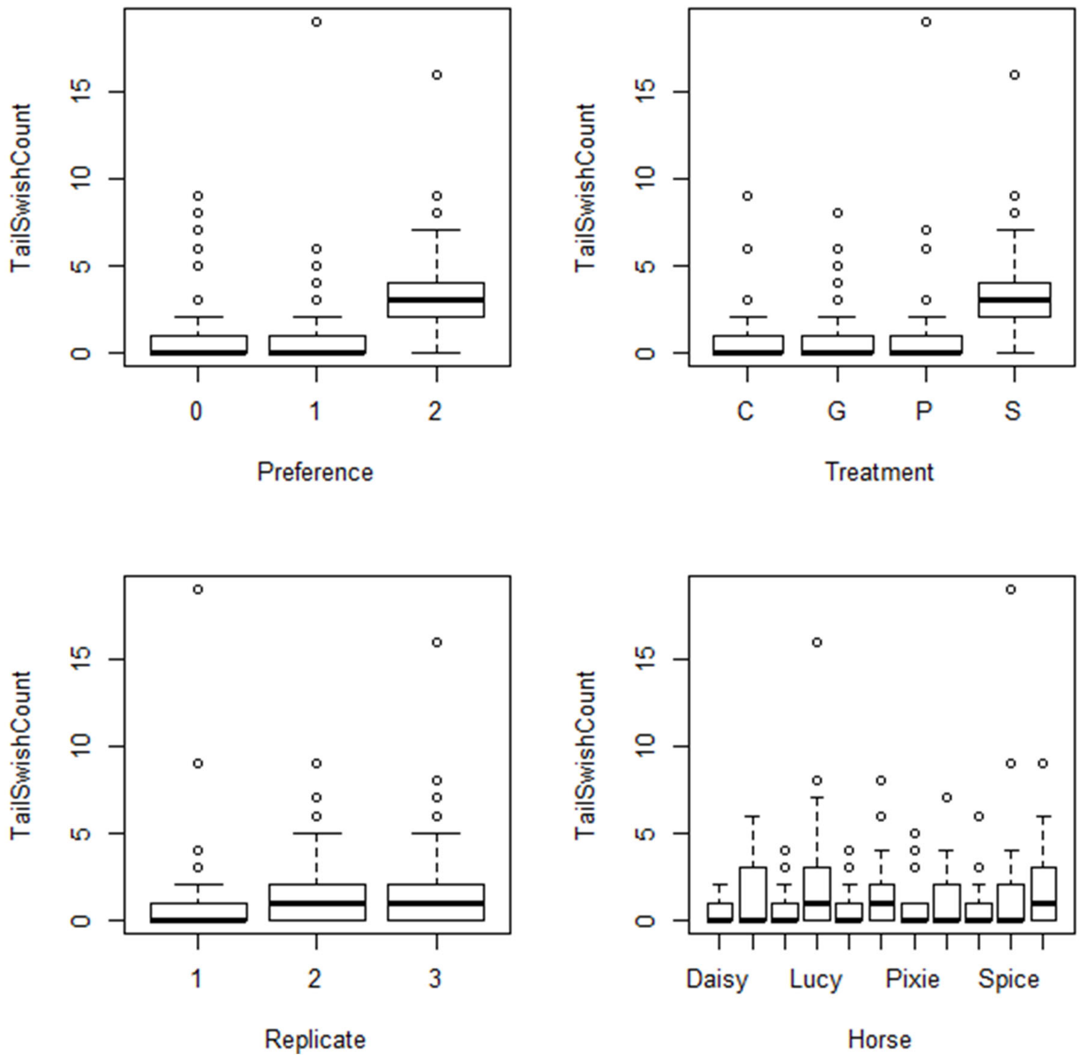


Figure A 19 Boxplots showing the tail swish count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray

Thrash

Table A 12 Two-way table showing the number of times a tail thrash did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Tail Thrash (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	59	7	66	11
	0	117	15	132	11
	2	25	8	33	24
Treatment	C	61	5	66	8
	G	57	9	66	14
	P	58	8	66	12
	S	25	8	33	24
Replicate	1	66	11	77	14
	2	74	3	77	4
	3	61	16	77	21
Horse	Daisy	15	6	21	29
	Frankie	21	0	21	0
	Grace	20	1	21	5
	Lucy	15	6	21	29
	Minty	20	1	21	5
	Phoenix	15	6	21	29
	Pixie Fae	18	3	21	14
	Semitone	20	1	21	5
	Shane	20	1	21	5
	Spice	17	4	21	19
	Whitney	20	1	21	5
Total		**201**	**30**	**231**	**13**

Tail Neutral

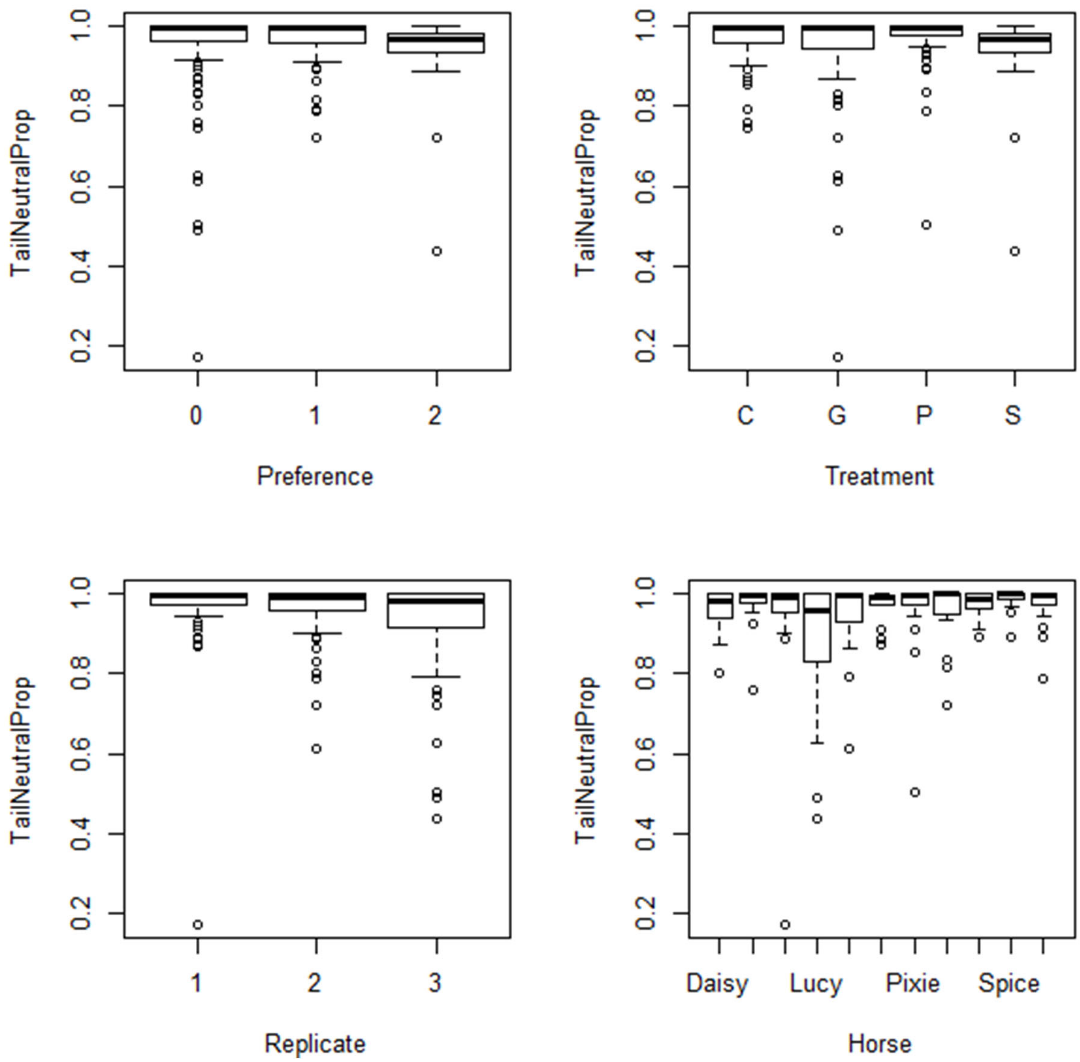


Figure A 20 Boxplots showing the tail neutral proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

A.5.5. Vocalisation

Snort

**Table A 13 Two-way table showing the number of times a snort did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.**

		Snort (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	58	8	66	12
	0	115	17	132	13
	2	25	8	33	24
Treatment	C	54	12	66	18
	G	58	8	66	12
	P	61	5	66	8
	S	25	8	33	24
Replicate	1	63	14	77	18
	2	71	6	77	8
	3	64	13	77	17
Horse	Daisy	18	3	21	14
	Frankie	18	3	21	14
	Grace	20	1	21	5
	Lucy	16	5	21	24
	Minty	20	1	21	5
	Phoenix	20	1	21	5
	Pixie Fae	12	9	21	43
	Semitone	21	0	21	0
	Shane	17	4	21	19
	Spice	19	2	21	10
	Whitney	17	4	21	19
Total		**198**	**33**	**231**	**14**

Whinny

Table A 14 Two-way table showing the number of times a whinny did and did not occur by preference, treatment, replicate, and horse. The % column shows the percentage of observations for which the behaviour was observed.

		Whinny (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	63	3	66	5
	0	125	7	132	5
	2	32	1	33	3
Treatment	C	62	4	66	6
	G	62	4	66	6
	P	64	2	66	3
	S	32	1	33	3
Replicate	1	71	6	77	8
	2	75	2	77	3
	3	74	3	77	4
Horse	Daisy	21	0	21	0
	Frankie	18	3	21	14
	Grace	21	0	21	0
	Lucy	18	3	21	14
	Minty	21	0	21	0
	Phoenix	21	0	21	0
	Pixie Fae	21	0	21	0
	Semitone	20	1	21	5
	Shane	17	4	21	19
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**220**	**11**	**231**	**5**

Nicker

Table A 15 Two-way table showing the number of times a nicker did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Nicker (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	132	0	132	0
	2	31	2	33	6
Treatment	C	65	1	66	2
	G	65	1	66	2
	P	66	0	66	0
	S	31	2	33	6
Replicate	1	75	2	77	3
	2	75	2	77	3
	3	77	0	77	0
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	20	1	21	5
	Minty	20	1	21	5
	Phoenix	21	0	21	0
	Pixie Fae	20	1	21	5
	Semitone	21	0	21	0
	Shane	20	1	21	5
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**227**	**4**	**231**	**2**

Any vocalisation

Table A 16 Two-way table showing the number of times any vocalisation did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Any vocal (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	54	12	66	18
	0	109	23	132	17
	2	24	9	33	27
Treatment	C	50	16	66	24
	G	54	12	66	18
	P	59	7	66	11
	S	24	9	33	27
Replicate	1	57	20	77	26
	2	69	8	77	10
	3	61	16	77	21
Horse	Daisy	18	3	21	14
	Frankie	15	6	21	29
	Grace	20	1	21	5
	Lucy	13	8	21	38
	Minty	19	2	21	10
	Phoenix	20	1	21	5
	Pixie Fae	12	9	21	43
	Semitone	20	1	21	5
	Shane	14	7	21	33
	Spice	19	2	21	10
	Whitney	17	4	21	19
Total		**187**	**44**	**231**	**19**

A.5.6. Elimination

Urination

Table A 17 Two-way table showing the number of times a urination did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Urination (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	0	0	0	0
	0	0	0	0	0
	2	0	0	0	0
Treatment	C	0	0	0	0
	G	0	0	0	0
	P	0	0	0	0
	S	0	0	0	0
Replicate	1	0	0	0	0
	2	0	0	0	0
	3	0	0	0	0
Horse	Daisy	0	0	0	0
	Frankie	0	0	0	0
	Grace	0	0	0	0
	Lucy	0	0	0	0
	Minty	0	0	0	0
	Phoenix	0	0	0	0
	Pixie Fae	0	0	0	0
	Semitone	0	0	0	0
	Shane	0	0	0	0
	Spice	0	0	0	0
	Whitney	0	0	0	0
Total		0	0	0	0

Defecation

Table A 18 Two-way table showing the number of times a defecation did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Defecation (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	61	5	66	8
	0	125	7	132	5
	2	31	2	33	6
Treatment	C	58	8	66	12
	G	63	3	66	5
	P	65	1	66	2
	S	31	2	33	6
Replicate	1	69	8	77	10
	2	74	3	77	4
	3	74	3	77	4
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	20	1	21	5
	Lucy	21	0	21	0
	Minty	16	5	21	24
	Phoenix	19	2	21	10
	Pixie Fae	19	2	21	10
	Semitone	20	1	21	5
	Shane	19	2	21	10
	Spice	21	0	21	0
	Whitney	20	1	21	5
Total		**217**	**14**	**231**	**6**

A.5.7. Analysis by preference

Table A 19 Results from models examining the effect of preference for a treatment on the occurrence, frequency, or proportion of time spent displaying various body behaviours.
Analysis is based on comparison to the preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Foreleg	Lift count [a]	Preferred	REF				3.1
		Intermediate	0.01 (0.07)	1.01 (0.88-1.16)	-	0.904	3.1
		Aversive	***0.79 (0.08)***	***2.21 (1.89-2.57)*** [d]	***-***	***<0.001****	***6.9***
	Strike [b^]	Preferred	REF				0.1%
		Intermediate	0.63 (0.83)	-	1.89 (0.37-9.53)	0.442	0.2%
		Aversive	1.14 (1.07)	-	3.14 (0.39-25.58)	0.285	0.3%
Hindleg	Lift count [a]	Preferred	REF				2.1
		Intermediate	0.11 (0.08)	1.12 (0.96-1.31)	-	0.159	2.4
		Aversive	***1.28 (0.08)***	***3.60 (3.06-4.23)***	***-***	***<0.001****	***7.7***
	Kick [b^]	Preferred	REF				0.6%
		Intermediate	0.66 (0.93)	-	1.94 (0.31-12.04)	0.476	1.2%
		Aversive	**1.92 (0.90)**	**-**	**6.81 (1.18-39.49)** [e]	**0.032***	**4.0%**
	Rest count [a]	Preferred	REF				2.0

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value#
		Intermediate	0.01 (0.10)	1.01 (0.83-1.22)	-	0.944	2.0
		Aversive	***0.69 (0.11)***	***1.99 (1.60-2.48)***	***-***	***<0.001****	***3.9***
Neck	Very low count [a]	Preferred	REF				2.4
		Intermediate	0.14 (0.08)	1.15 (0.99-1.34)	-	0.066	2.8
		Aversive	***0.34 (0.10)***	***1.41 (1.16-1.72)***	***-***	***<0.001****	***3.4***
	Low time [c]	Intercept (Preferred)	0.11 (0.03)	-	-	<0.001	11.1%
		Intermediate	0.01 (0.01)	-	-	0.666	12.1%
		Aversive	**-0.05 (0.02) [f]**	**-**	**-**	**0.015***	**6.3%**
	Neutral time [c]	Intercept (Preferred)	0.60 (0.05)	-	-	<0.001	59.7
		Intermediate	-0.01 (0.03)	-	-	0.689	58.6
		Aversive	***-0.14 (0.04)***	***-***	***-***	***<0.001****	***46.0***
	High time [c]	Intercept (Preferred)	0.25 (0.05)	-	-	<0.001	24.6%
		Intermediate	0.001 (0.03)	-	-	0.982	24.7%
		Aversive	***0.19 (0.04)***	***-***	***-***	***<0.001****	***43.6%***
	Very high count [a]	Preferred	REF				0.2
		Intermediate	***-0.52 (0.18)***	***0.59 (0.42-0.84)***	***-***	***0.004****	***0.1***
		Aversive	***0.58 (0.19)***	***1.78 (1.23-2.59)***	***-***	***0.002****	***0.3***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Tail	Lift [b^]	Preferred	REF				24.4%
		Intermediate	0.17 (0.36)	-	1.18 (0.59-2.38)	0.633	27.7%
		Aversive	0.17 (0.50)	-	1.18 (0.45-3.14)	0.733	27.7%
	Swish count [a]	Preferred	REF				0.8
		Intermediate	-0.25 (0.15)	0.78 (0.58-1.04)	-	0.095	0.6
		Aversive	***1.24 (0.15)***	***3.47 (2.60-4.63)***	***-***	***<0.001****	***2.9***
	Thrash [b^]	Preferred	REF				4.1%
		Intermediate	0.1 (0.56)	-	1.11 (0.37-3.31)	0.854	4.6%
		Aversive	**1.38 (0.69)**	**-**	**3.98 (1.03-15.33)**	**0.045***	**14.7%**
Vocalisation	Snort [b^]	Preferred	REF				9.5%
		Intermediate	0.08 (0.48)	-	1.08 (0.42-2.75)	0.875	10.3%
		Aversive	0.93 (0.59)	-	2.54 (0.80-8.05)	0.112	21.1%
	Whinny [b^]	Preferred	REF				1.6%
		Intermediate	0.18 (0.74)	-	1.19 (0.28-5.09)	0.810	1.9%
		Aversive	-0.45 (1.21)	-	0.64 (0.06-6.87)	0.709	1.0%
	Any [b^]	Preferred	REF				15.6%
		Intermediate	-0.06 (0.40)	-	0.95 (0.43-2.09)	0.889	14.8%
		Aversive	-0.58 (0.53)	-	1.78 (0.364-4.99)	0.272	24.8%

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Elimination	Defecation [b^]	Preferred	REF				5.6%
		Intermediate	-0.4 (0.62)	-	0.67 (0.20-2.27)	0.521	3.8%
		Aversive	-0.25 (0.89)	-	0.78 (0.14-4.41)	0.777	4.4%

* $p \leq 0.05$
^ Caution with interpretation as large SE.
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the Preferred treatment, the rate of foreleg lift behaviour was 2.21 (95%CI 1.89 - 2.57) times greater during the Aversive treatment and this was highly statistically significant ($p < 0.001$). Foreleg lift was predicted to occur 6.9 times during the three-minute Aversive treatment.
[e] Compared to the Preferred treatment, the odds of a hindleg kick occurring was 6.81 (95%CI 1.18 - 39.49) times higher during the Aversive treatment and this was statistically significant ($p = 0.032$). There was a 4.0% predicted probability of occurrence of a hindleg kick during the Aversive treatment. Due to a large SE of the beta, the confidence interval around the odds ratio is wide, and caution with interpretation is advised.
[f] Compared to the Preferred treatment, the proportion of time that horses had their neck in the low/very low position was 0.05 (SE 0.02) less during the Aversive treatment, this was statistically significant ($p = 0.015$). The predicted percentage of time that horses had their neck in the low position was 11% for the Preferred treatment and 6% for the Aversive treatment.

A.5.8. Analysis by treatment

Table A 20 Results from models examining the effect of treatment on the occurrence, frequency or proportion of time spent displaying various body behaviours.
Analysis is based on comparison to the no-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Foreleg	Lift count [a]	No-stimulus	REF				3.1
		Groom	-0.05 (0.08)	0.95 (0.81-1.12)	-	0.538	3.0
		Person	0.07 (0.08)	1.07 (0.92-1.25)	-	0.388	3.3
		Spray	***0.79 (0.08)***	***2.21 (1.90-2.58)***	***-***	***<0.001****	***6.9***
	Strike [b^]	No-stimulus	REF				0.1%
		Groom	-1.74 (1.04)	-	0.18 (0.02-1.34)	0.094	0.03%
		Person	-0.73 (0.88)	-	0.48 (0.09-2.73)	0.409	0.1%
		Spray	-0.73 (0.80)	-	0.48 (0.10-2.32)	0.362	0.1%
Hindleg	Lift count [a]	No-stimulus	REF				2.3
		Groom	0.16 (0.09)	1.18 (0.99-1.39)	-	0.063	2.7
		Person	**-0.20 (0.10)**	**0.82 (0.68-0.99)** [d]	**-**	**0.036***	**1.9**
		Spray	***1.20 (0.08)***	***3.33 (2.84-3.90)***	***-***	***<0.001****	***7.7***
	Kick [b^]	No-stimulus	REF				0.3%
		Groom	2.16 (1.26)	-	8.67 (0.73-102.39)	0.086	2.1%

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
		Person	0.77 (1.42)	-	2.15 (0.13-34.49)	0.588	0.5%
		Spray	**2.72 (1.21)**	**-**	**15.12 (1.41-162.61)**	**0.025***	**3.6%**
	Rest count [a]	No-stimulus	REF				1.8
		Groom	0.09 (0.11)	1.10 (0.88-1.37)	-	0.423	2.0
		Person	0.18 (0.11)	1.20 (0.96-1.50)	-	0.104	2.2
		Spray	***0.78 (0.11)***	***2.18 (1.74-2.73)***	***-***	***<0.001****	***3.9***
Neck	Very low count [a]	No-stimulus	REF				2.5
		Groom	**0.18 (0.09)**	**1.20 (1.01-1.42)**	**-**	**0.037***	**2.9**
		Person	0.04 (0.09)	1.04 (0.87-1.24)	-	0.688	2.6
		Spray	***0.32 (0.10)***	***1.38 (1.13-1.68)***	***-***	***0.001****	***3.4***
	Low time [c]	Intercept (No-stimulus)	0.09 (0.03)	-	-	0.004	9.0%
		Groom	0.03 (0.02)	-	-	0.094	11.6%
		Person	***0.05 (0.02)***	***-***	***-***	***0.003****	***13.8%***
		Spray	-0.03 (0.02)	-	-	0.158	6.3%
	Neutral time [c]	Intercept (No-stimulus)	0.60 (0.05)	-	-	<0.001	59.9
		Groom	-0.04 (0.03)	-	-	0.237	56.3
		Person	0.01 (0.03)	-	-	0.772	60.8
		Spray	***-0.14 (0.04)***	***-***	***-***	***<0.001****	***46.0***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	High time [c]	Intercept (No-stimulus)	0.27 (0.05)	-	-	<0.001	26.5%
		Groom	0.01 (0.03)	-	-	0.807	27.2%
		Person	**-0.06 (0.03)**	**-**	**-**	**0.031***	**20.1%**
		Spray	***0.17 (0.04)*** [e]	***-***	***-***	***<0.001****	***43.5%***
	Very high count [a]	No-stimulus	REF				0.1
		Groom	**0.46 (0.22)**	**1.58 (1.02-2.45)**	**-**	**0.039***	**0.2**
		Person	0.24 (0.23)	1.27 (0.81-2.01)	-	0.296	0.1
		Spray	***1.14 (0.22)***	***3.14 (2.03-4.86)***	***-***	***<0.001****	***0.3***
Tail	Lift [b^]	No-stimulus	REF				34.3%
		Groom	-0.32 (0.40)	-	0.73 (0.33-1.58)	0.422	27.5%
		Person	**-0.87 (0.42)**	**-**	**0.42 (0.18-0.96)**	**0.040***	**18.0%**
		Spray	-0.32 (0.49)	-	0.73 (0.28-1.90)	0.514	27.5%
	Swish count [a]	No-stimulus	REF				0.5
		Groom	0.34 (0.19)	1.40 (0.96-2.03)	-	0.078	0.8
		Person	**0.41 (0.19)**	**1.51 (1.04-2.17)**	**-**	**0.029***	**0.8**
		Spray	***1.67 (0.17)***	***5.30 (3.79-7.40)***	***-***	***<0.001****	***2.9***

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Thrash [b^]	No-stimulus	REF				2.5%
		Groom	0.87 (0.68)	-	2.39 (0.63-9.06)	0.200	5.8%
		Person	0.69 (0.69)	-	1.99 (0.52-7.66)	0.319	4.9%
		Spray	**1.88 (0.75)**	**-**	**6.54 (1.51-28.39)**	**0.012***	**14.4%**
Vocalisation	Snort [b^]	No-stimulus	REF				14.8%
		Groom	-0.53 (0.52)	-	0.59 (0.21-1.64)	0.311	9.3%
		Person	-1.09 (0.59)	-	0.34 (0.11-1.07)	0.066	5.5%
		Spray	-0.41 (0.55)	-	1.51 (0.51-4.46)	0.454	20.8%
	Whinny [b^]	No-stimulus	REF				2.2%
		Groom	0.00 (0.77)	-	1.00 (0.22-4.54)	1.00	2.2%
		Person	-0.79 (0.92)	-	0.45 (0.07-2.76)	0.391	1.0%
		Spray	-0.79 (1.18)	-	0.45 (0.04-4.59)	0.504	1.0%
	Any [b^]	No-stimulus	REF				21.4%
		Groom	-0.40 (0.45)	-	0.67 (0.28-1.60)	0.366	15.5%
		Person	**-1.08 (0.51)**	**-**	**0.34 (0.13-0.92)** [f]	**0.033***	**8.5%**
		Spray	0.18 (0.51)	-	1.20 (0.44-3.24)	0.726	24.6%

Body part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Elimination	Defecation [b^]	No-stimulus	REF				8.9%
		Groom	-1.14 (0.73)	-	0.32 (0.08-1.33)	0.118	3.0%
		Person	**-2.30 (1.10)**	**-**	**0.10 (0.01-0.86)**	**0.036***	**1.0%**
		Spray	-0.82 (0.85)	-	0.44 (0.08-2.33)	0.336	4.1%

* p ≤ 0.05
^ Caution with interpretation as large SE.
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the No-stimulus treatment, the rate of hindleg lift behaviour was 0.82 (95%CI 0.68 - 0.99) times less during the Person treatment, this was statistically significant (p = 0.036). Hindleg lift was predicted to occur 1.9 times during the three-minute Person treatment.
[e] Compared to the No-stimulus treatment, the proportion of time that horses had their neck in the high position was 0.17 (SE 0.04) more during the Spray treatment, this was statistically significant (p < 0.001). The predicted percentage of time that horses had their neck high/very high position was 27% for the No-stimulus treatment and 44% for the Spray treatment.
[f] Compared to the No-stimulus treatment, the odds of any vocalisation occurring was 0.34 (95%CI 0.13 - 0.92) times lower during the Person treatment and this was statistically significant (p = 0.033). There was an 8.5% predicted probability of occurrence of a vocalisation during the Person treatment.

A.6. Chapter 9 Supplementary material

A.6.1. Head

Head shaking

Table A 21 Two-way table showing the number of times a head shake did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Head shaking (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	51	15	66	23
	0	91	41	132	31
	2	23	10	33	30
Treatment	C	52	14	66	21
	G	39	27	66	41
	P	51	15	66	23
	S	23	10	33	30
Replicate	1	46	31	77	40
	2	57	20	77	26
	3	62	15	77	19
Horse	Daisy	14	7	21	33
	Frankie	18	3	21	14
	Grace	15	6	21	29
	Lucy	10	11	21	52
	Minty	15	6	21	29
	Phoenix	13	8	21	38
	Pixie Fae	10	11	21	52
	Semitone	16	5	21	24
	Shane	19	2	21	10
	Spice	15	6	21	29

		Head shaking (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
	Whitney	20	1	21	5
Total		**165**	**66**	**231**	**29**

Head tossing

Table A 22 Two-way table showing the number of times a head toss did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Head tossing (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	57	9	66	14
	0	111	21	132	16
	2	24	9	33	27
Treatment	C	53	13	66	20
	G	58	8	66	12
	P	57	9	66	14
	S	24	9	33	27
Replicate	1	61	16	77	21
	2	66	11	77	14
	3	65	12	77	16
Horse	Daisy	14	7	21	33
	Frankie	20	1	21	5
	Grace	20	1	21	5
	Lucy	16	5	21	24
	Minty	20	1	21	5
	Phoenix	18	3	21	14
	Pixie Fae	17	4	21	19
	Semitone	21	0	21	0
	Shane	11	10	21	48
	Spice	15	6	21	29
	Whitney	20	1	21	5
Total		**192**	**39**	**231**	**17**

Head weaving

Table A 23 Two-way table showing the number of times a head weave did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Head weaving (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	132	0	132	0
	2	33	0	33	0
Treatment	C	65	1	66	2
	G	66	0	66	0
	P	65	1	66	2
	S	33	0	33	0
Replicate	1	76	1	77	1
	2	77	0	77	0
	3	76	1	77	1
Horse	Daisy	21	0	21	0
	Frankie	20	1	21	5
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	20	1	21	5
	Phoenix	21	0	21	0
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**229**	**2**	**231**	**1**

A.6.2. Ears

Left ear forward

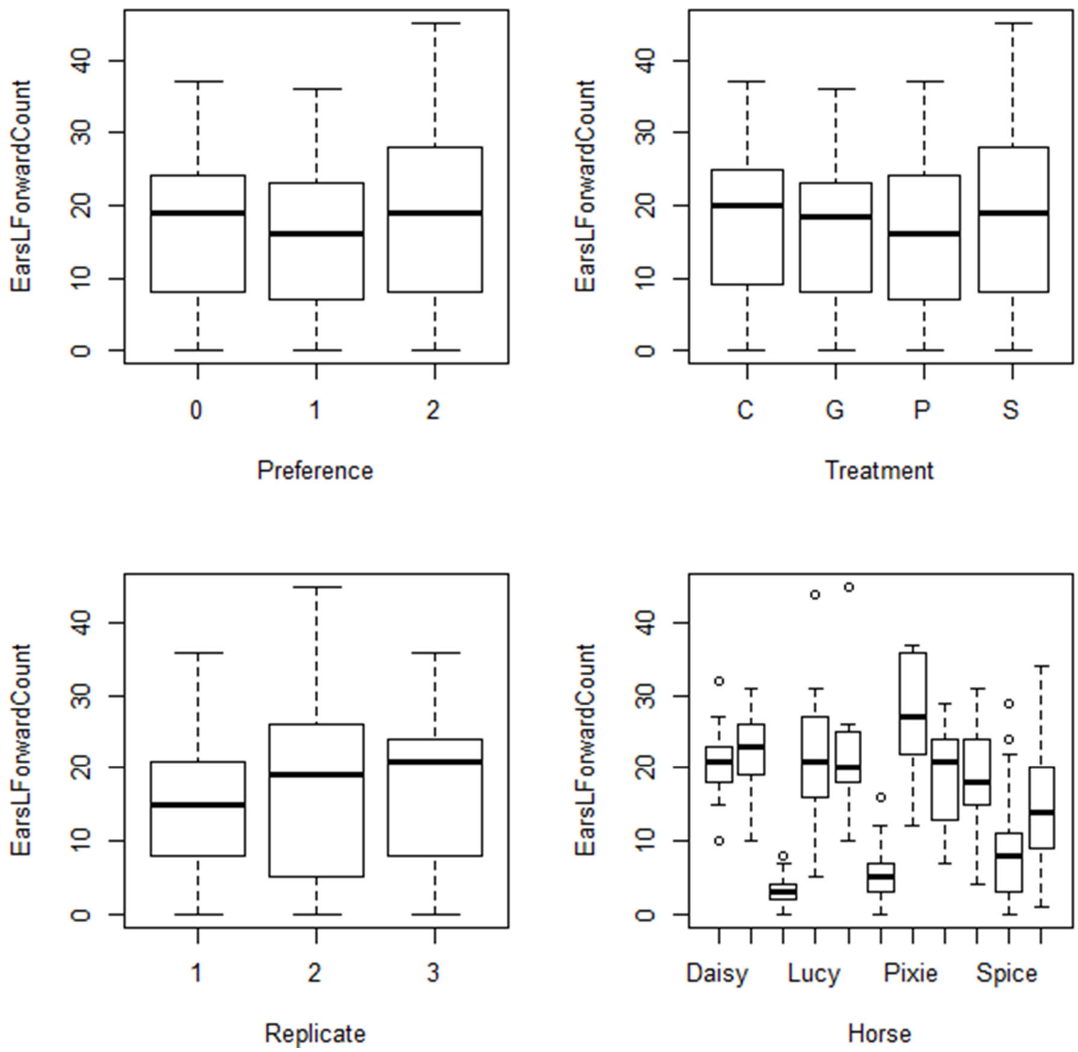


Figure A 21 Boxplots showing the left ear forward count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

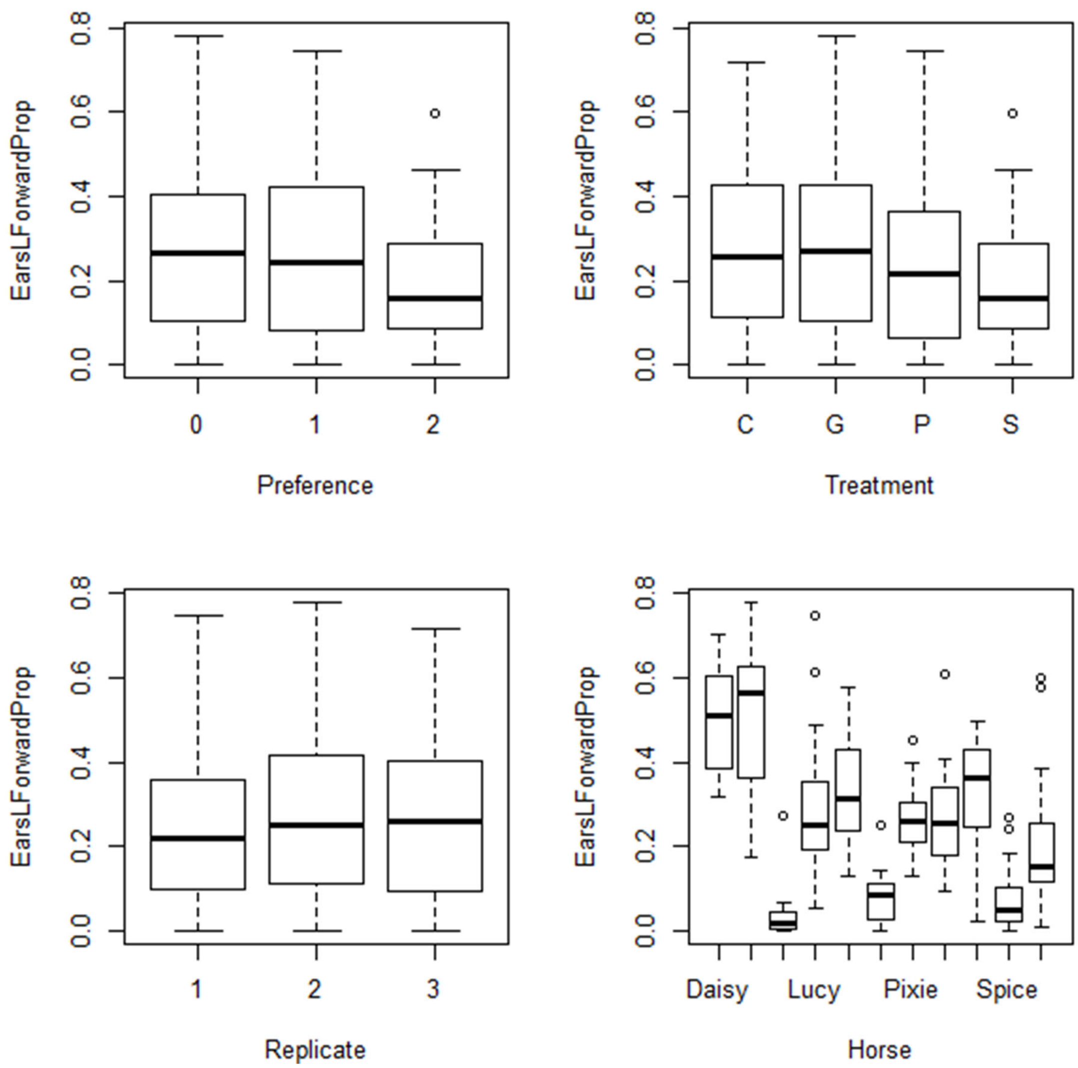


Figure A 22 Boxplots showing the left ear forward proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Right ear forward

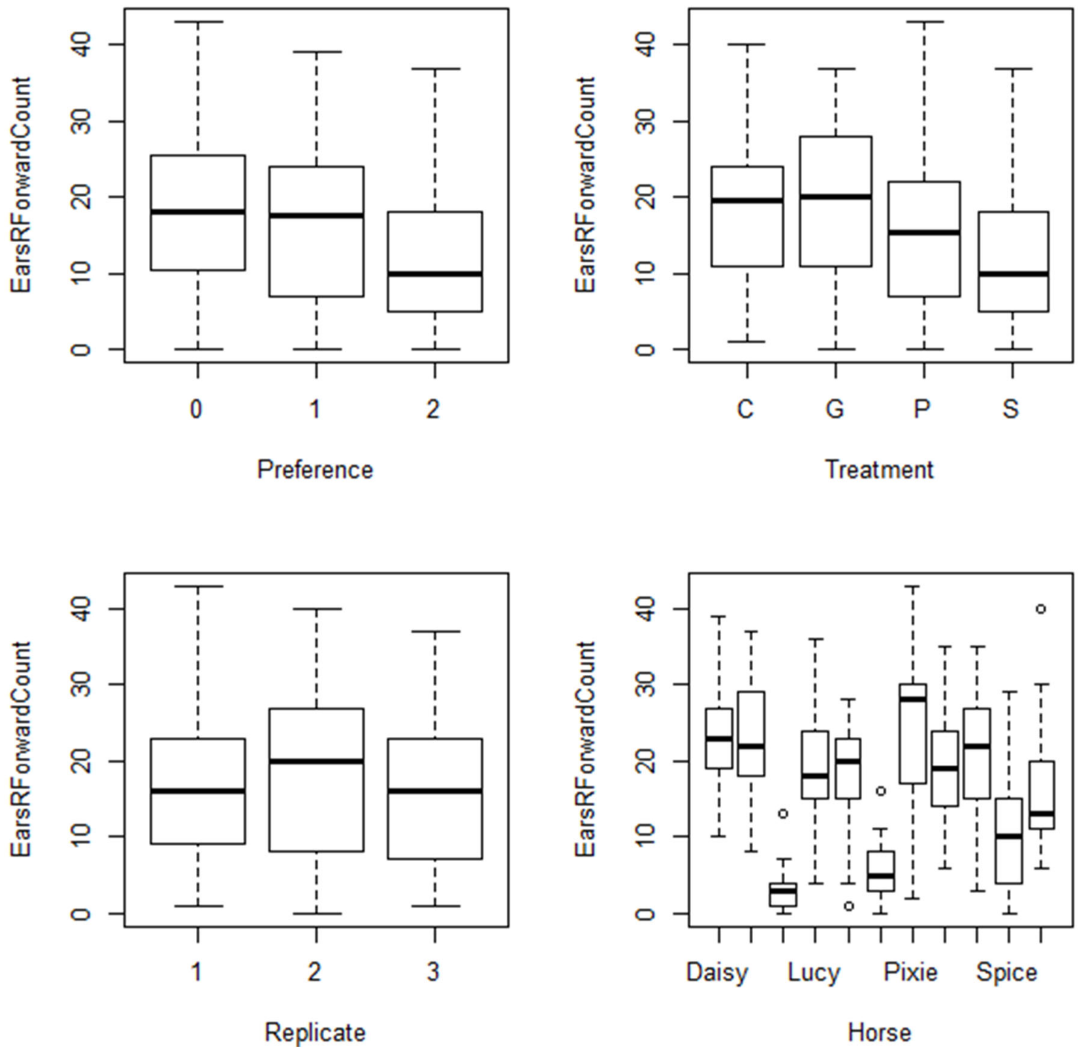


Figure A 23 Boxplots showing the right ear forward count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

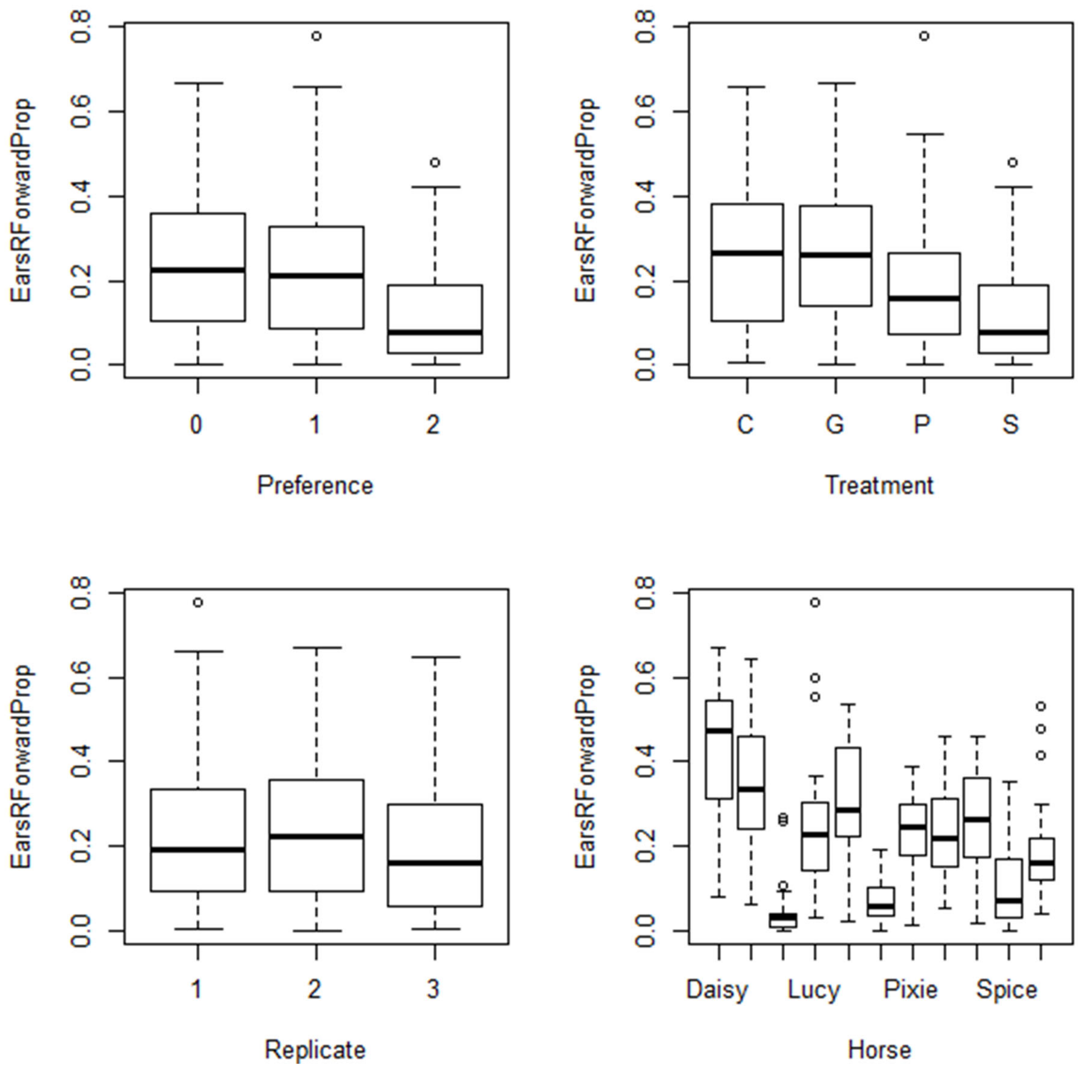


Figure A 24 Boxplots showing the right ear forward proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Left ear side

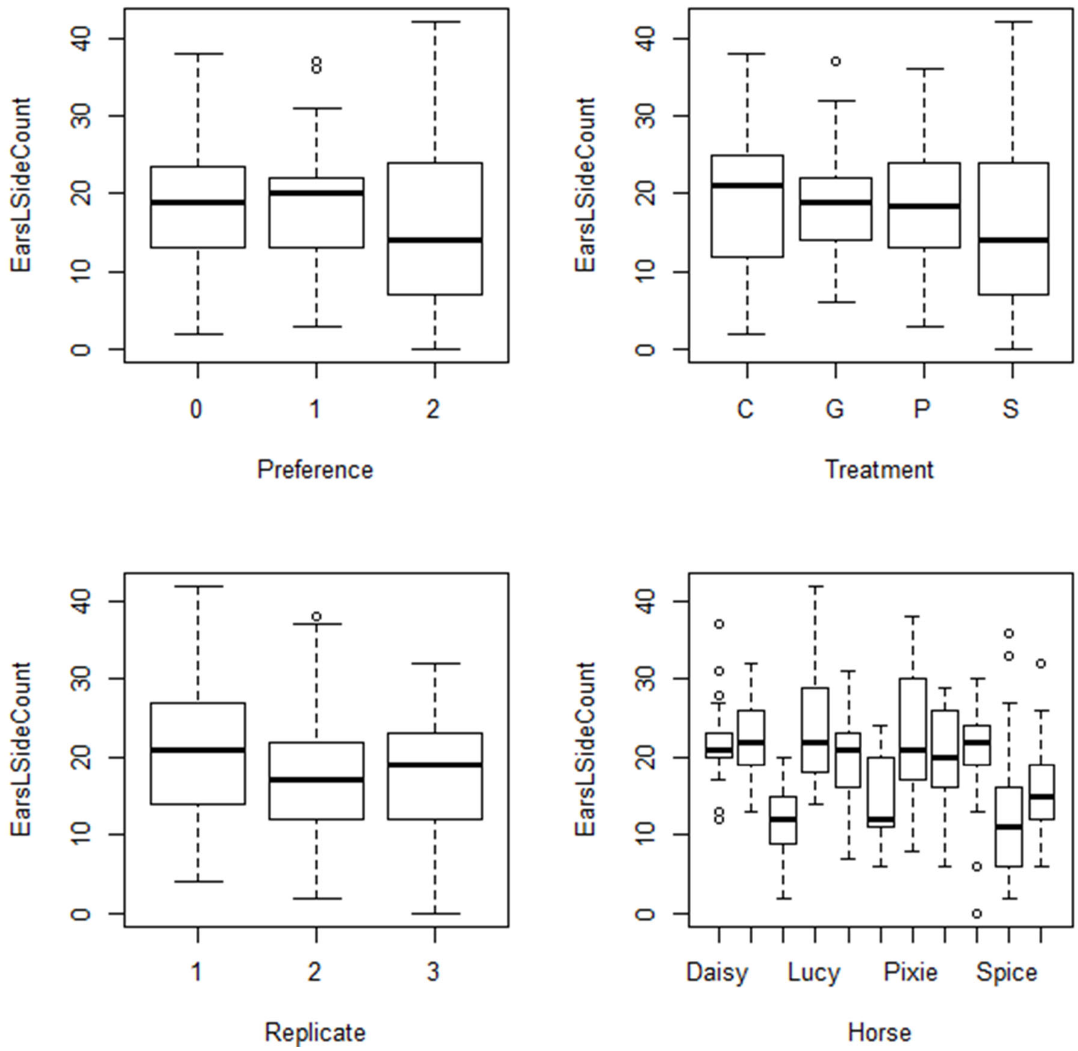


Figure A 25 Boxplots showing the left ear side count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

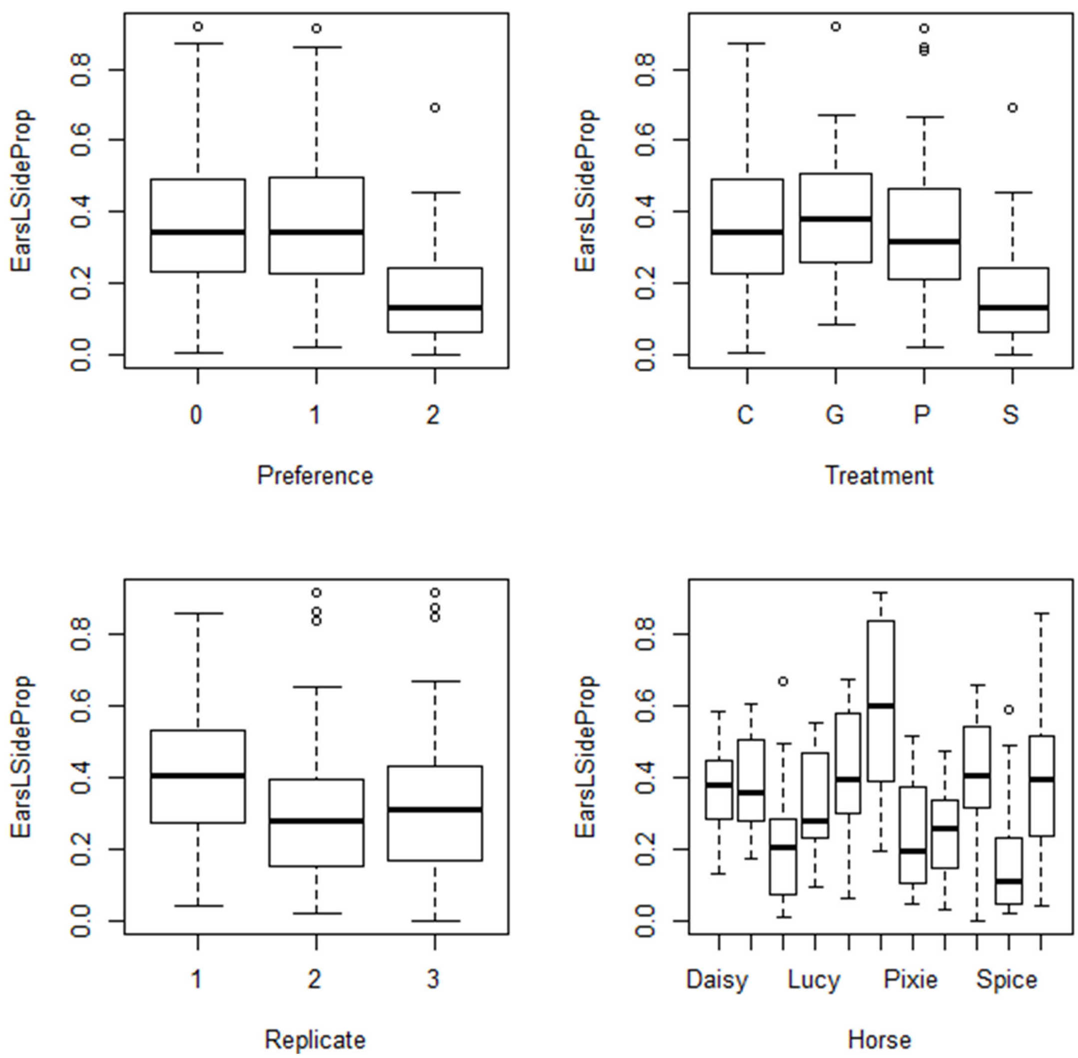


Figure A 26   Boxplots showing the left ear side proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Right ear side

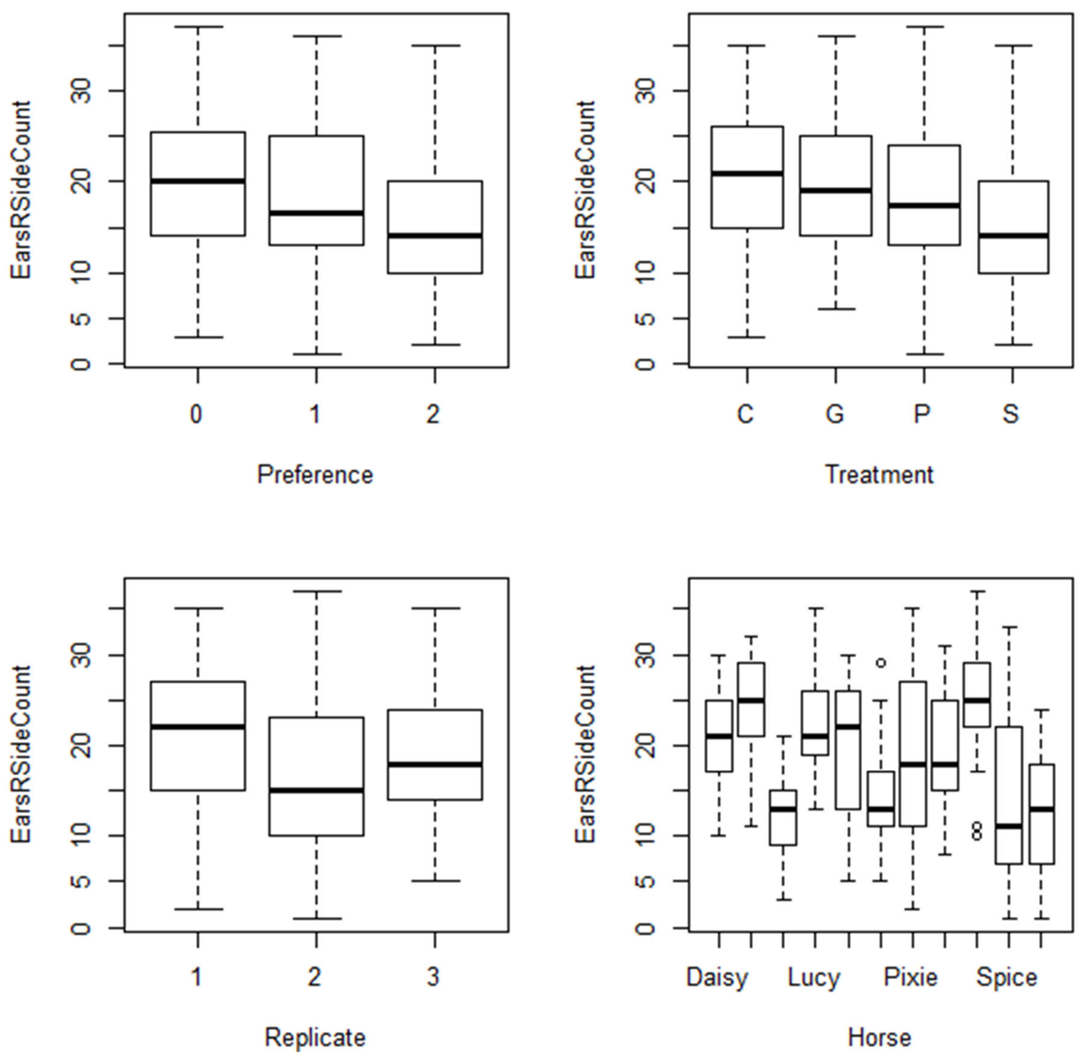


Figure A 27 Boxplots showing the right ear side count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

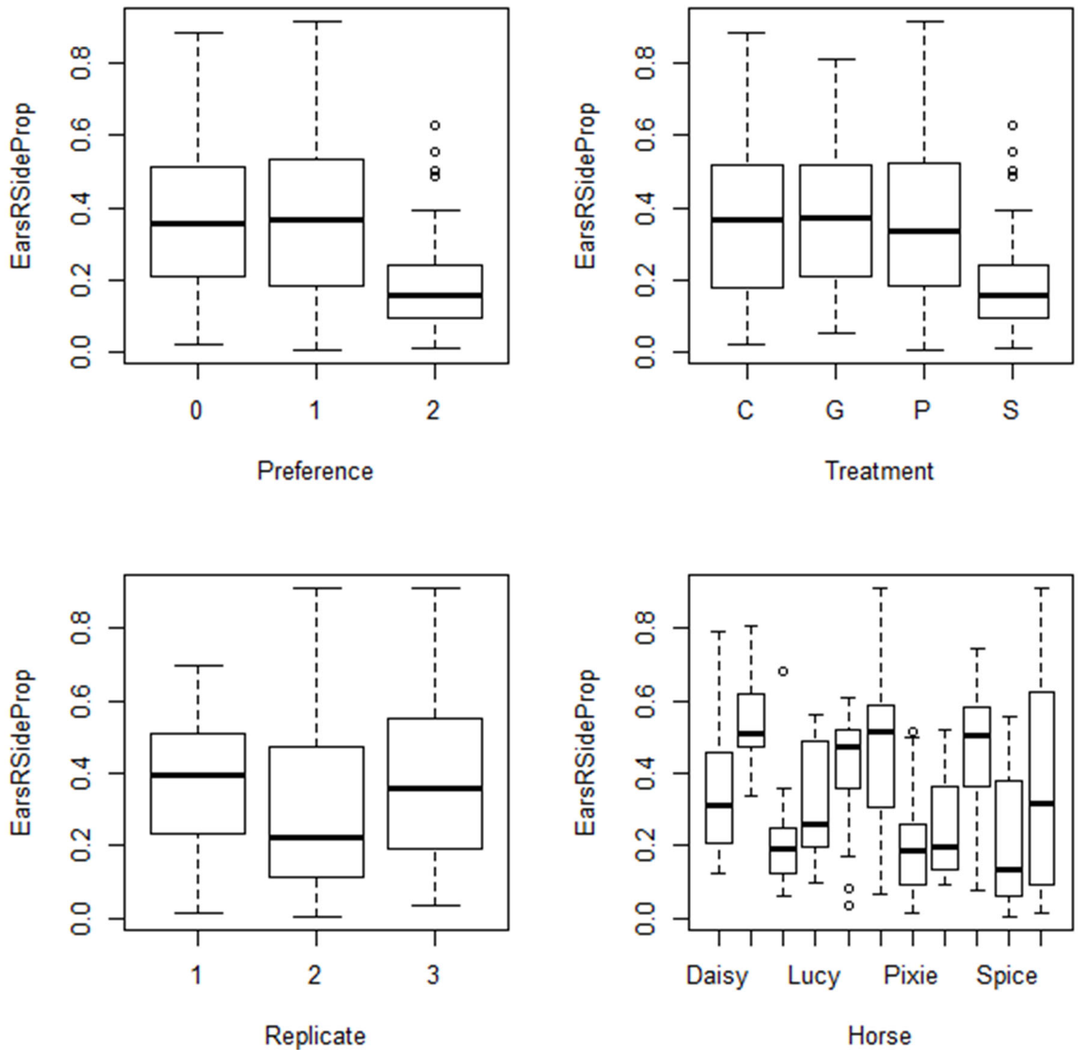


Figure A 28 Boxplots showing the right ear side proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Left ear back

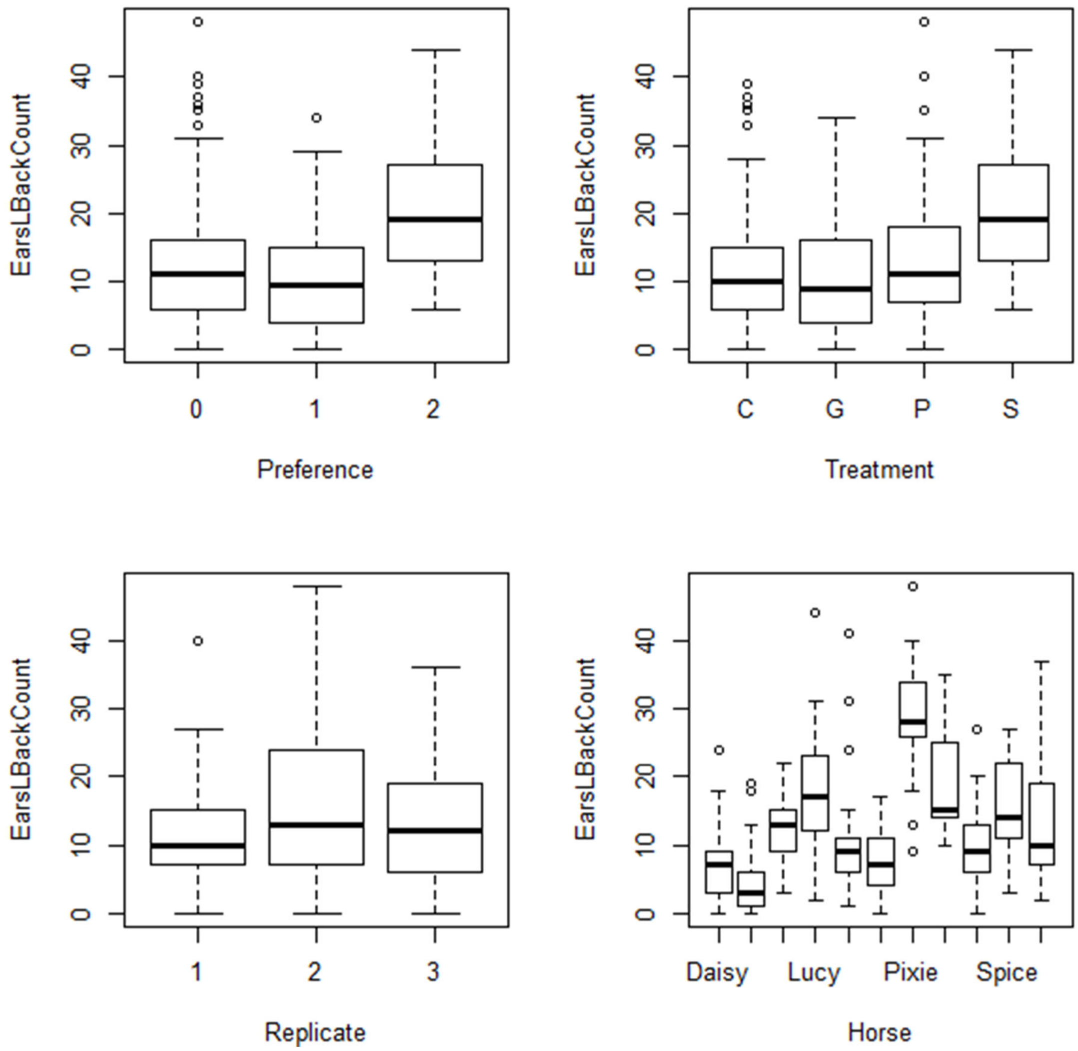


Figure A 29 Boxplots showing the left ear back count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

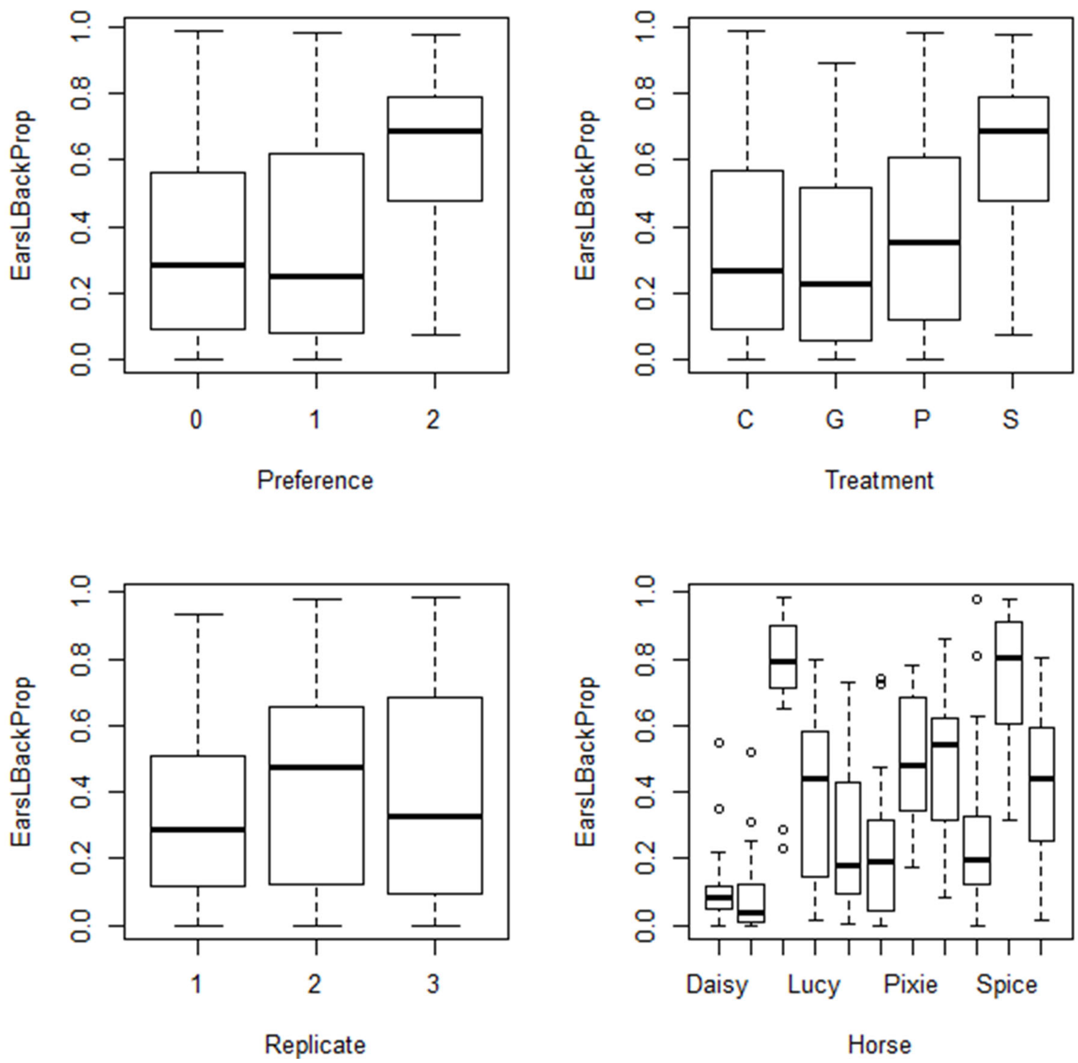


Figure A 30 Boxplots showing the left ear back proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Right ear back

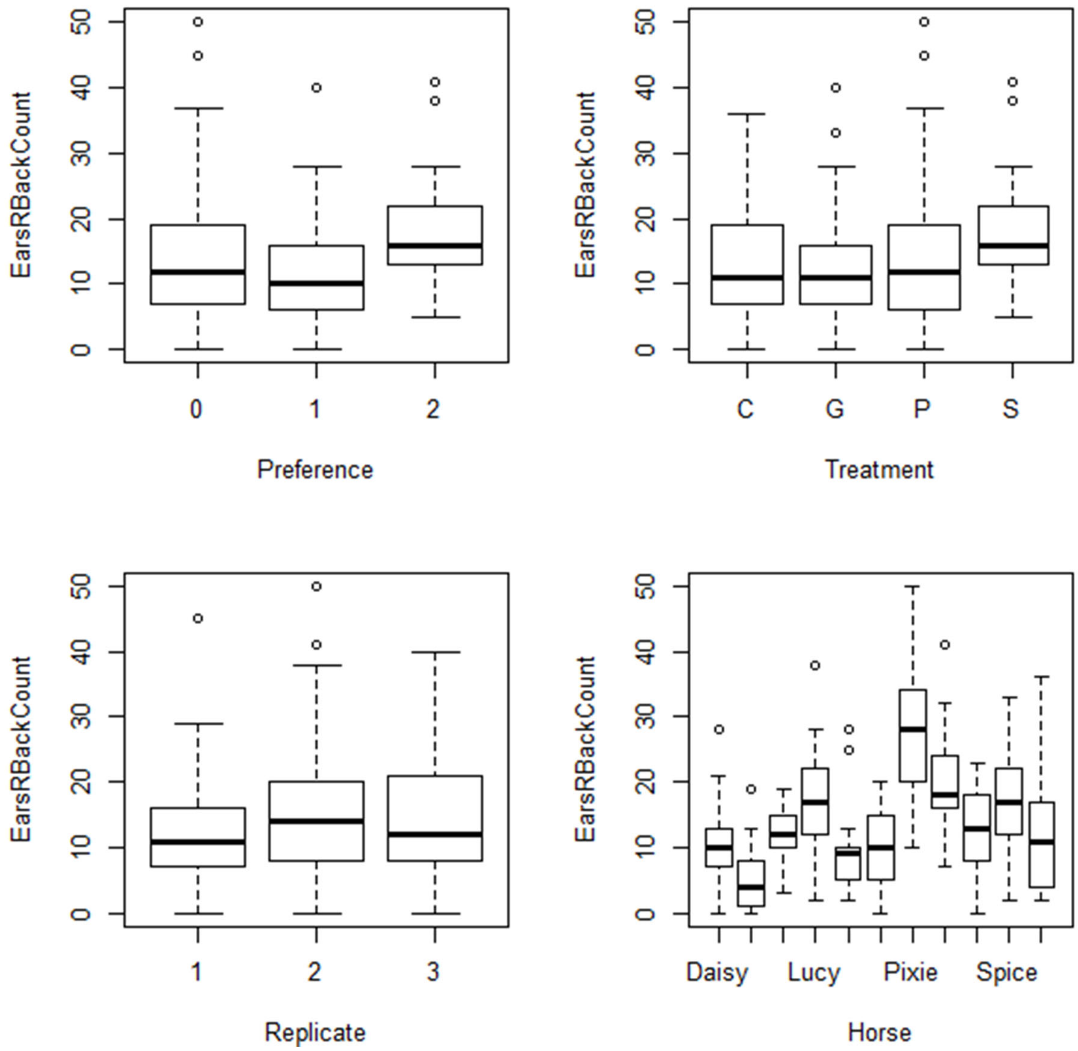


Figure A 31 Boxplots showing the right ear back count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

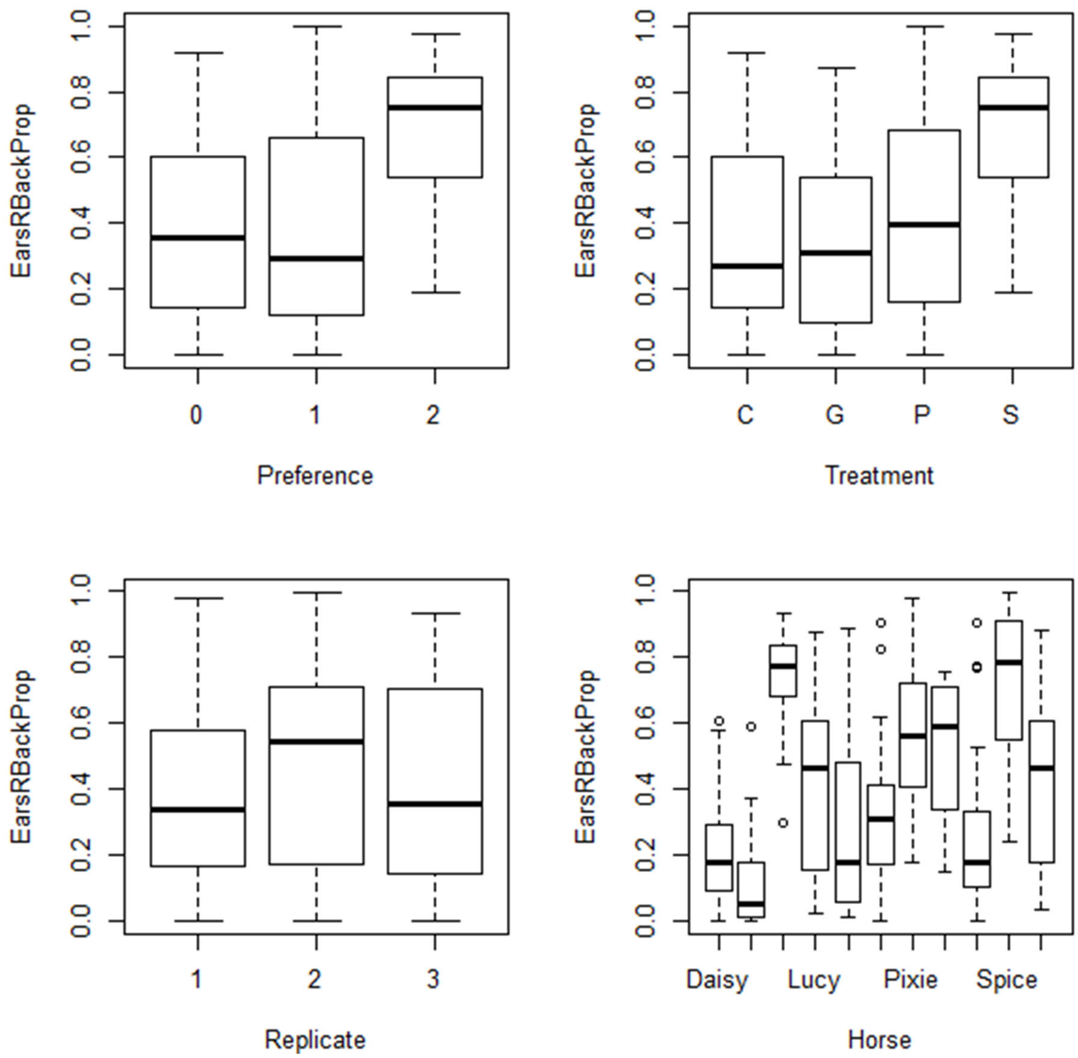


Figure A 32 Boxplots showing the right ear back proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Left ear flat

Table A 24 Two-way table showing the number of times a left ear flat did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Left ear flat (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	62	4	66	6
	0	124	8	132	6
	2	29	4	33	12
Treatment	65	65	1	66	2
	G	61	5	66	8
	P	60	6	66	9
	S	29	4	33	12
Replicate	1	70	7	77	9
	2	72	5	77	6
	3	73	4	77	5
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	20	1	21	5
	Lucy	21	0	21	0
	Minty	20	1	21	5
	Phoenix	11	10	21	48
	Pixie Fae	19	2	21	10
	Semitone	21	0	21	0
	Shane	20	1	21	5
	Spice	20	1	21	5
	Whitney	21	0	21	0
Total		**215**	**16**	**231**	**7**

Right ear flat

Table A 25 Two-way table showing the number of times a right ear flat did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Right ear flat (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	62	4	66	6
	0	124	8	132	6
	2	28	5	33	15
Treatment	65	65	1	66	2
	G	61	5	66	8
	P	60	6	66	9
	S	28	5	33	15
Replicate	1	69	8	77	10
	2	72	5	77	6
	3	73	4	77	5
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	20	1	21	5
	Lucy	21	0	21	0
	Minty	20	1	21	5
	Phoenix	11	10	21	48
	Pixie Fae	19	2	21	10
	Semitone	20	1	21	5
	Shane	20	1	21	5
	Spice	20	1	21	5
	Whitney	21	0	21	0
Total		**214**	**17**	**231**	**7**

Left ear flicking

Table A 26 Two-way table showing the number of times a left ear flick did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Left ear flicking (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**% Rear**
Preference	1	66	0	66	0
	0	131	1	132	1
	2	33	0	33	0
Treatment	C	65	1	66	2
	G	66	0	66	0
	P	66	0	66	0
	S	33	0	33	0
Replicate	1	76	1	77	1
	2	77	0	77	0
	3	77	0	77	0
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	21	0	21	0
	Phoenix	21	0	21	0
	Pixie Fae	20	1	21	5
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**230**	**1**	**231**	**0**

Right ear flicking

Table A 27 Two-way table showing the number of times a right ear flick did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Right ear flicking (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**% Rear**
Preference	1	66	0	66	0
	0	132	0	132	0
	2	32	1	33	3
Treatment	C	66	0	66	0
	G	66	0	66	0
	P	66	0	66	0
	S	32	1	33	3
Replicate	1	77	0	77	0
	2	77	0	77	0
	3	76	1	77	1
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	21	0	21	0
	Phoenix	20	1	21	5
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**230**	**1**	**231**	**0**

A.6.3. Eyes

Blink

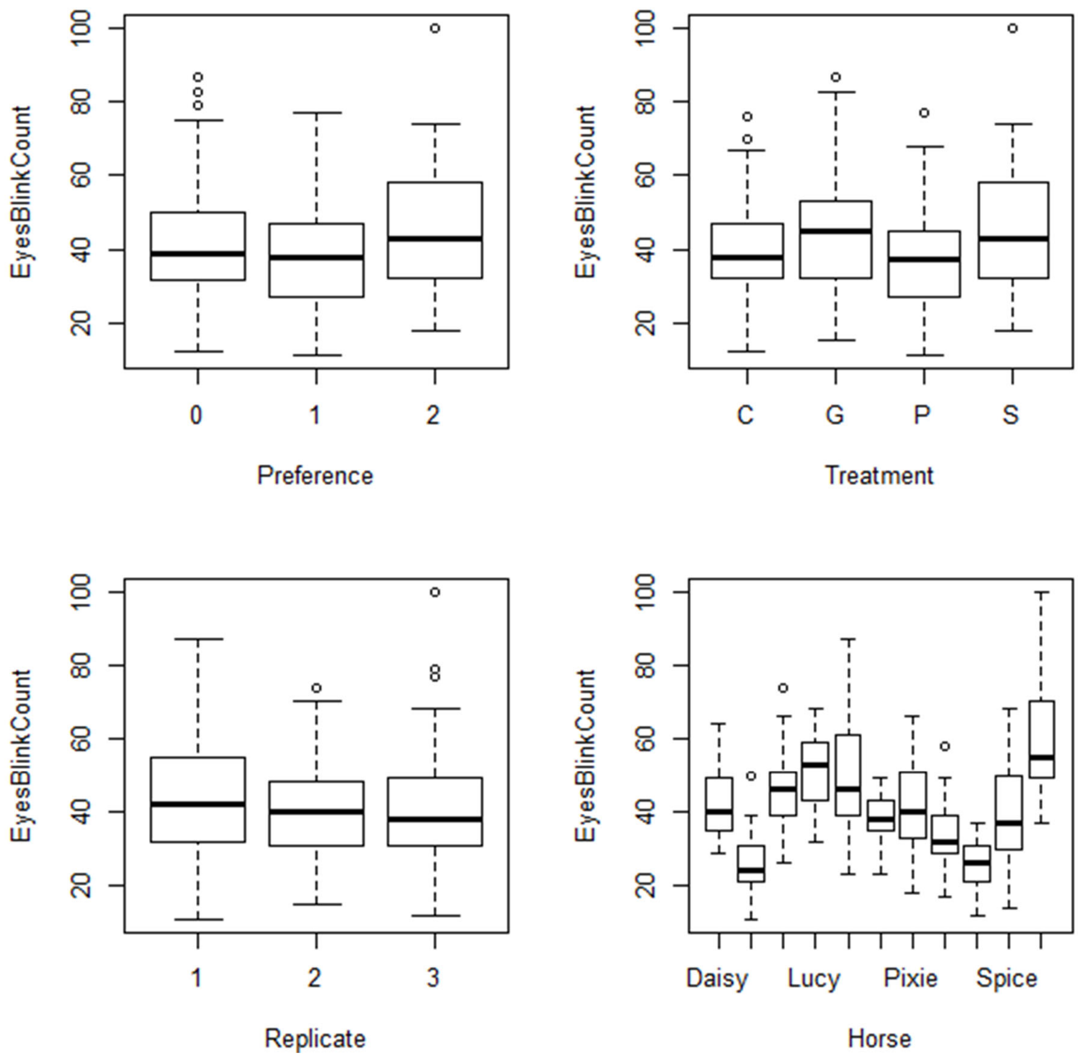


Figure A 33 Boxplots showing the blink count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Sclera visible

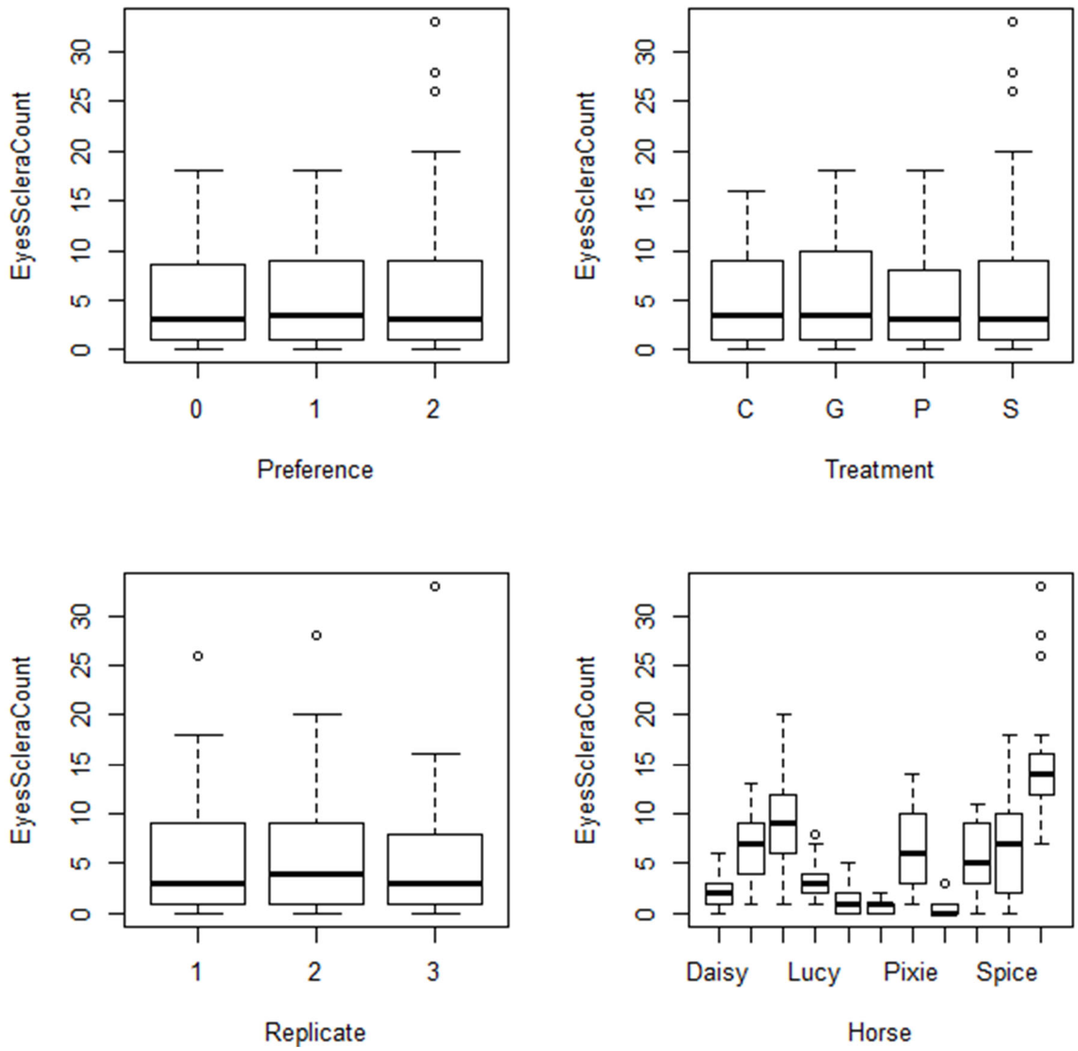


Figure A 34 Boxplots showing the sclera visible count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Small eye aperture

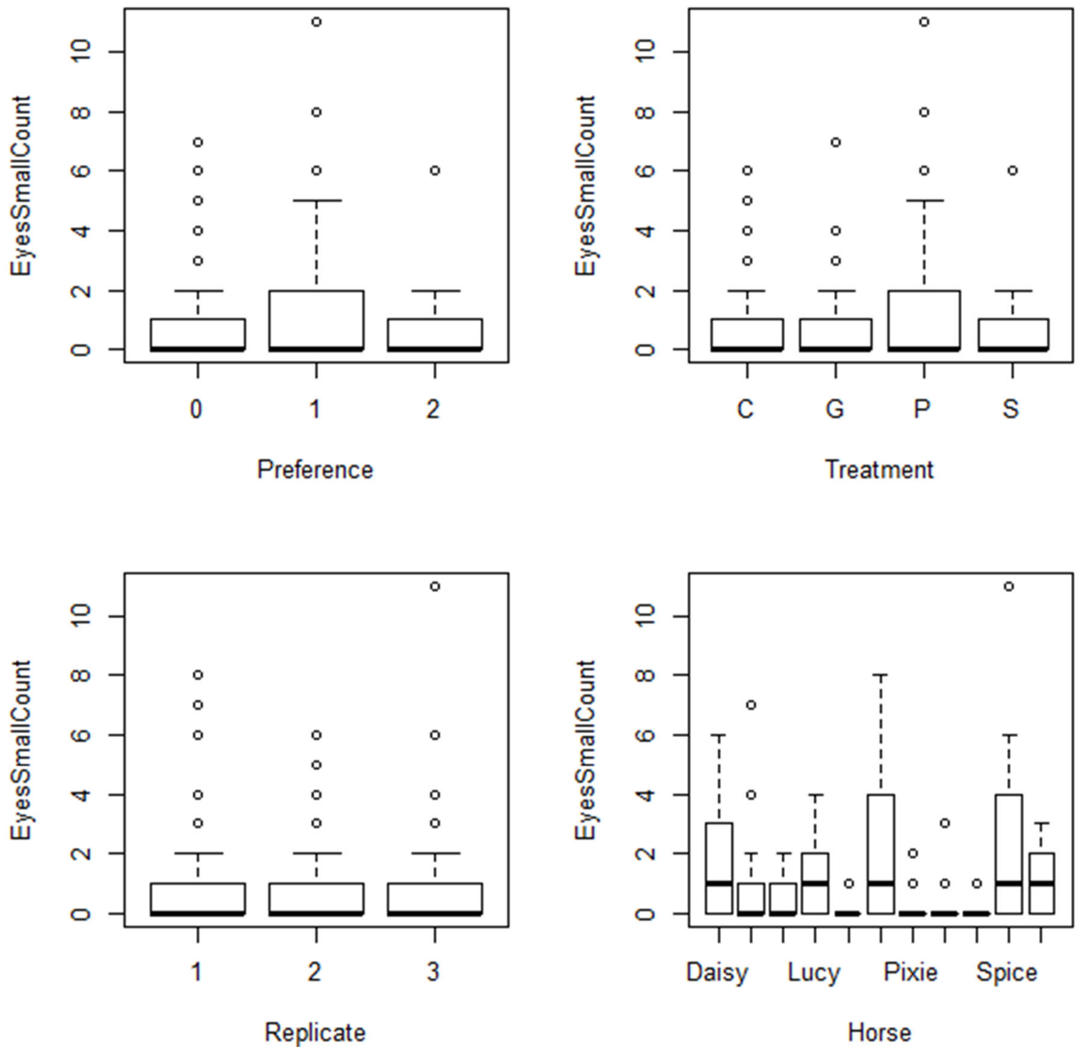


Figure A 35 Boxplots showing the small eye aperture count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Neutral eye aperture

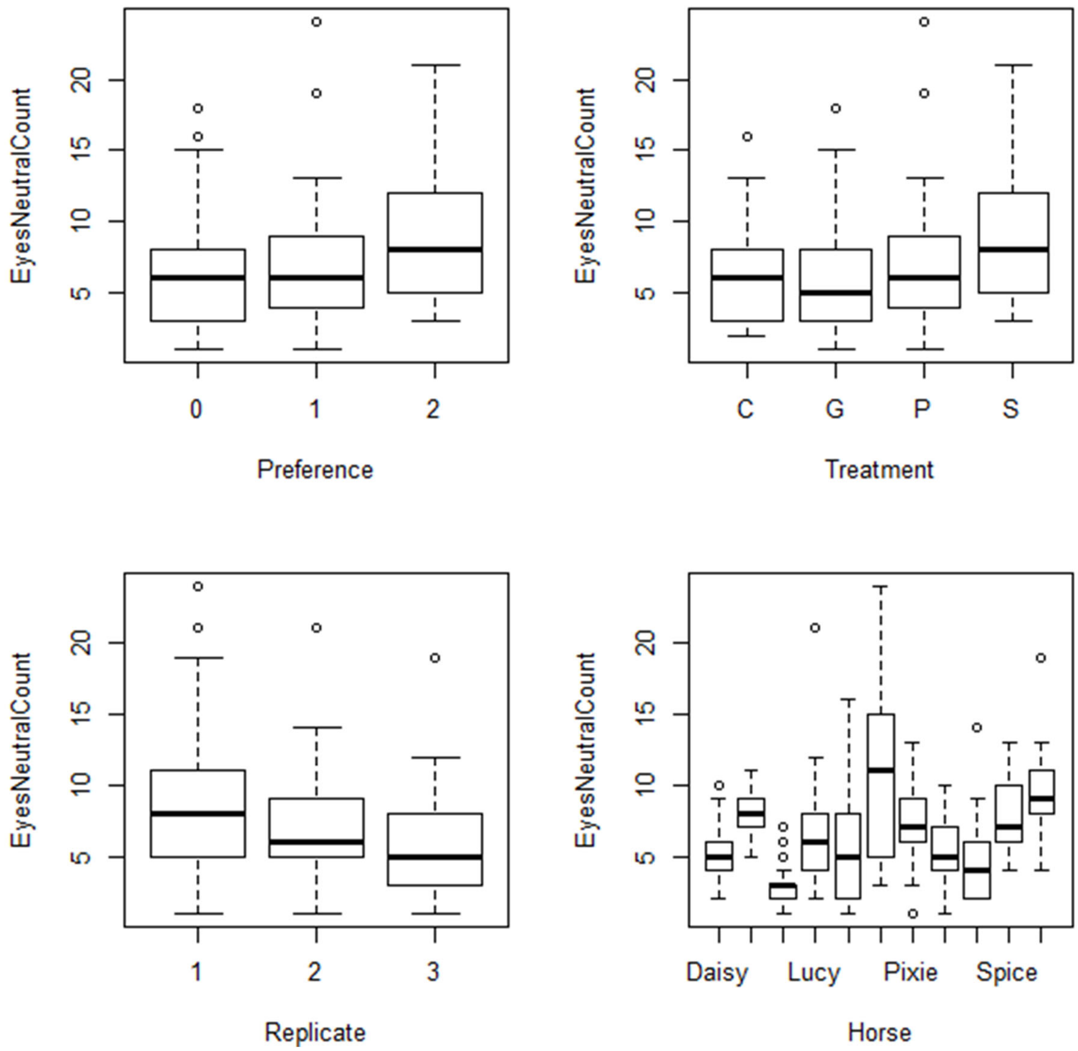


Figure A 36   Boxplots showing the neutral eye aperture count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Large eye aperture

Table A 28 Two-way table showing the number of times a large eye aperture did and did not occur by preference, treatment, replicate, and horse. The % column shows the percentage of observations for which the behaviour was observed.

		Large eye aperture (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	61	5	66	8
	0	124	8	132	6
	2	20	13	33	39
Treatment	C	63	3	66	5
	G	62	4	66	6
	P	60	6	66	9
	S	20	13	33	39
Replicate	1	67	10	77	13
	2	71	6	77	8
	3	67	10	77	13
Horse	Daisy	20	1	21	5
	Frankie	20	1	21	5
	Grace	20	1	21	5
	Lucy	19	2	21	10
	Minty	20	1	21	5
	Phoenix	12	9	21	43
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	18	3	21	14
	Spice	20	1	21	5
	Whitney	14	7	21	33
Total		**205**	**26**	**231**	**11**

Eyebrow wrinkle

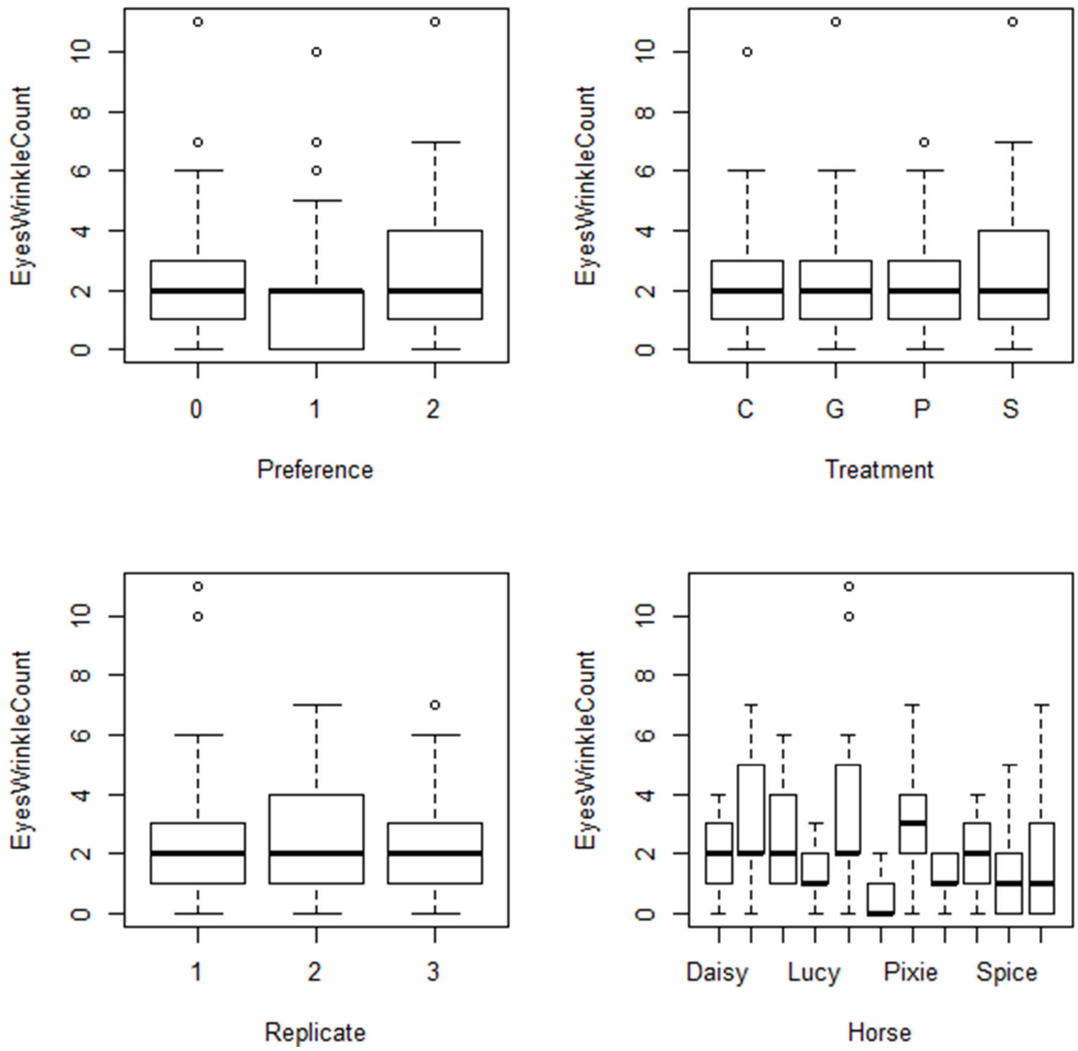


Figure A 37 Boxplots showing the eyebrow wrinkle count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Eyebrow angle

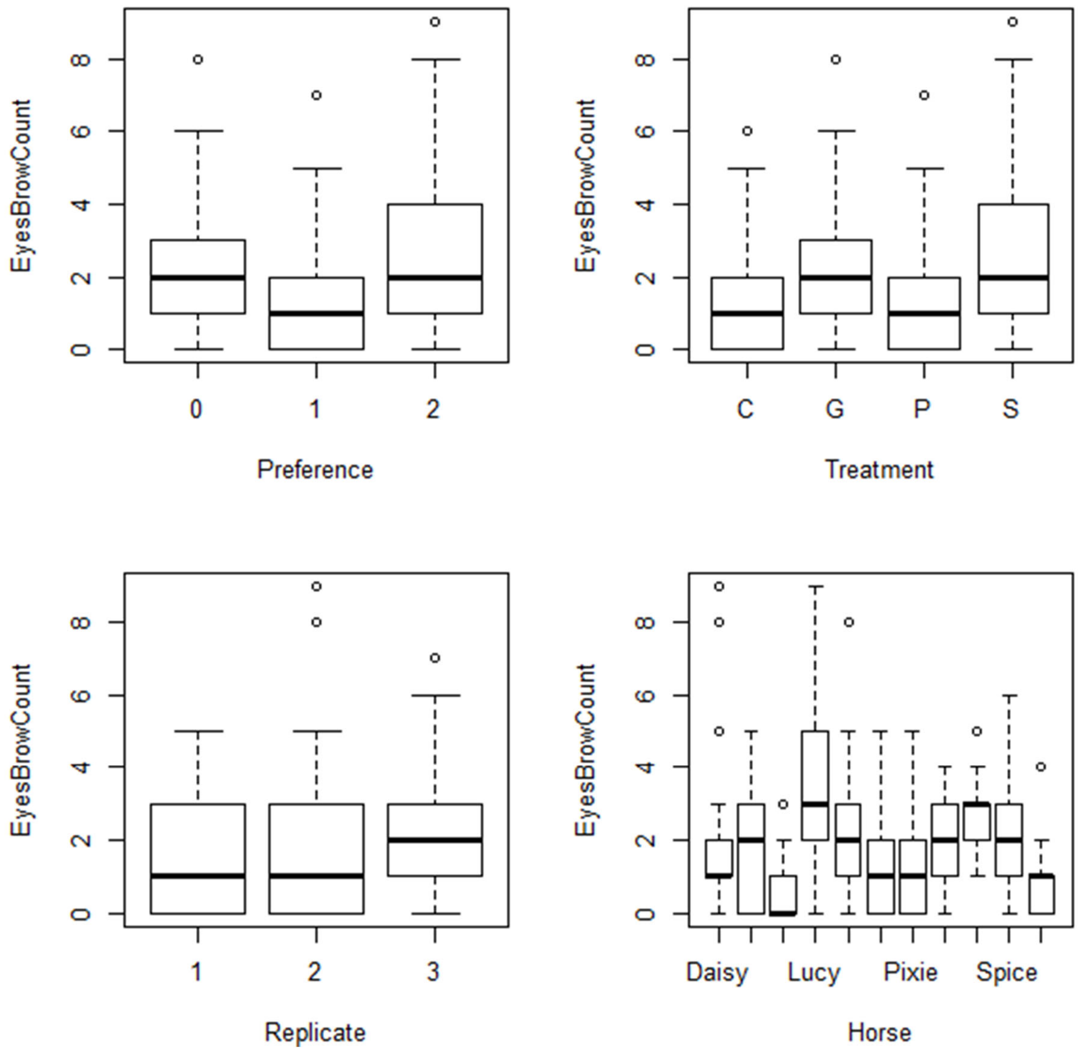


Figure A 38 Boxplots showing the eyebrow angle count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

A.6.4. Mouth

Contracted lips

Table A 29 Two-way table showing the number of times a contracted lips did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Contracted lips (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	58	8	66	12
	0	126	6	132	5
	2	16	17	33	52
Treatment	C	63	3	66	5
	G	64	2	66	3
	P	57	9	66	14
	S	16	17	33	52
Replicate	1	64	13	77	17
	2	67	10	77	13
	3	69	8	77	10
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	17	4	21	19
	Lucy	19	2	21	10
	Minty	18	3	21	14
	Phoenix	14	7	21	33
	Pixie Fae	20	1	21	5
	Semitone	20	1	21	5
	Shane	18	3	21	14
	Spice	16	5	21	24
	Whitney	16	5	21	24
Total		**200**	**31**	**231**	**13**

Drooping lip

Table A 30 Two-way table showing the number of times a drooping lip did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Drooping lip (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	37	29	66	44
	0	75	57	132	43
	2	25	8	33	24
Treatment	C	37	29	66	44
	G	34	32	66	48
	P	41	25	66	38
	S	25	8	33	24
Replicate	1	57	20	77	26
	2	43	34	77	44
	3	37	40	77	52
Horse	Daisy	13	8	21	38
	Frankie	17	4	21	19
	Grace	6	15	21	71
	Lucy	21	0	21	0
	Minty	5	16	21	76
	Phoenix	21	0	21	0
	Pixie Fae	11	10	21	48
	Semitone	3	18	21	86
	Shane	19	2	21	10
	Spice	7	14	21	67
	Whitney	14	7	21	33
Total		**137**	**94**	**231**	**41**

Lip licking and chewing

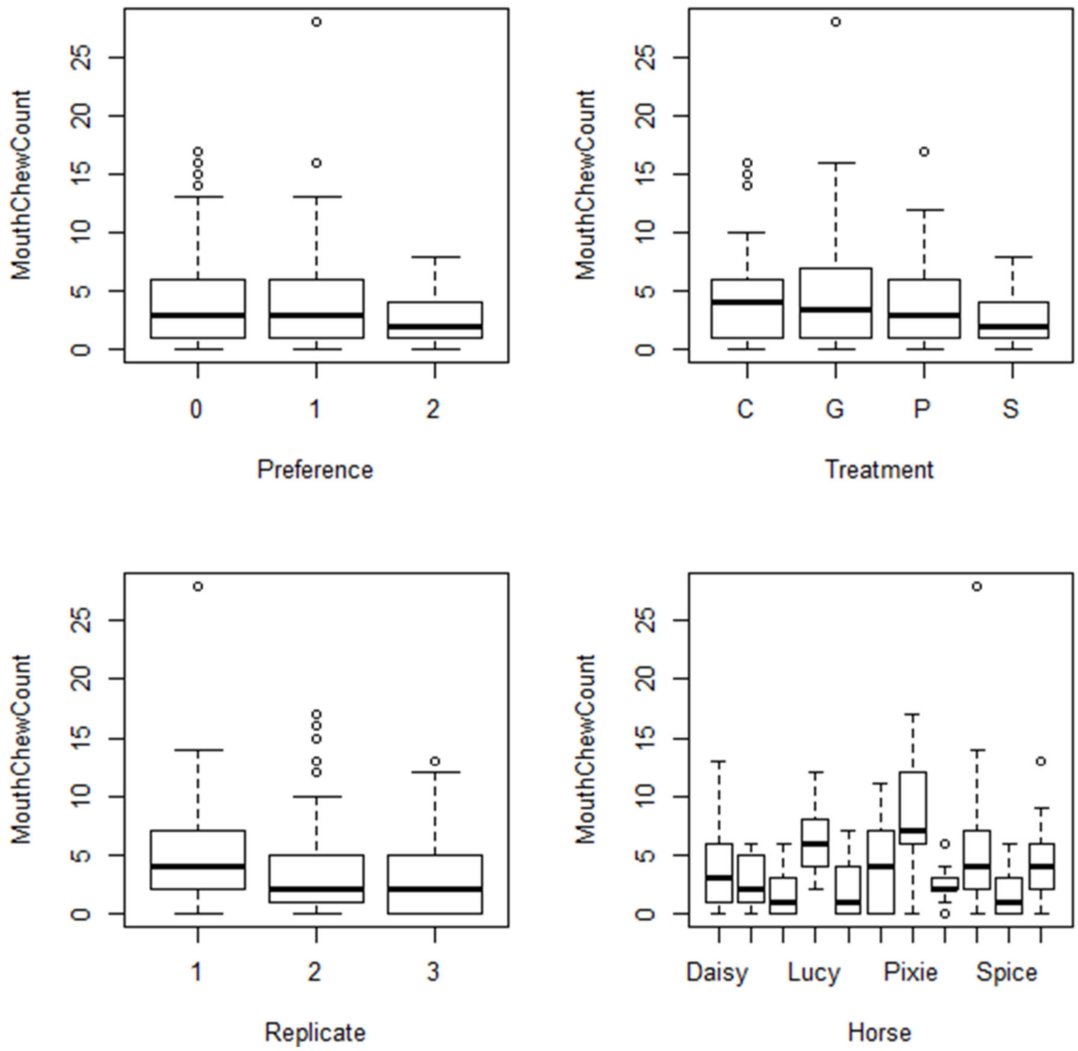


Figure A 39 Boxplots showing the lip licking and chewing count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Face pulling

Table A 31 Two-way table showing the number of times face pulling did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Face pulling (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	53	13	66	20
	0	111	21	132	16
	2	29	4	33	12
Treatment	C	55	11	66	17
	G	52	14	66	21
	P	57	9	66	14
	S	29	4	33	12
Replicate	1	61	16	77	21
	2	64	13	77	17
	3	68	9	77	12
Horse	Daisy	19	2	21	10
	Frankie	20	1	21	5
	Grace	21	0	21	0
	Lucy	0	21	21	100
	Minty	21	0	21	0
	Phoenix	21	0	21	0
	Pixie Fae	10	11	21	52
	Semitone	20	1	21	5
	Shane	20	1	21	5
	Spice	20	1	21	5
	Whitney	21	0	21	0
Total		**193**	**38**	**231**	**16**

Oral investigation

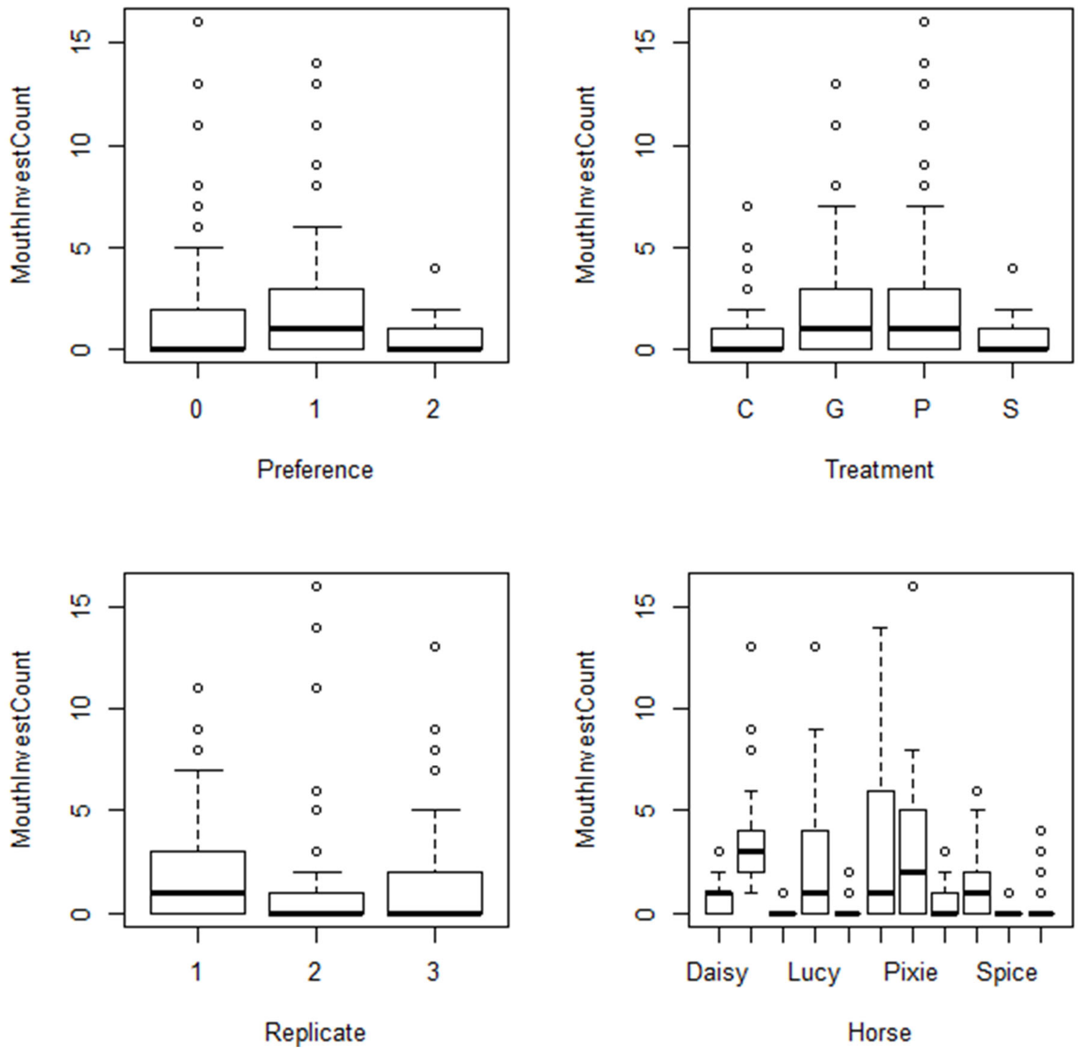


Figure A 40 Boxplots showing the oral investigation count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Yawn

Table A 32 Two-way table showing the number of times a yawn did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Yawn (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	65	1	66	2
	0	118	14	132	11
	2	33	0	33	0
Treatment	C	63	3	66	5
	G	55	11	66	17
	P	65	1	66	2
	S	33	0	33	0
Replicate	1	73	4	77	5
	2	68	9	77	12
	3	75	2	77	3
Horse	Daisy	19	2	21	10
	Frankie	18	3	21	14
	Grace	20	1	21	5
	Lucy	21	0	21	0
	Minty	21	0	21	0
	Phoenix	18	3	21	14
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	18	3	21	14
	Spice	20	1	21	5
	Whitney	19	2	21	10
Total		**216**	**15**	**231**	**6**

Eating

Table A 33 Two-way table showing the number of times eating did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Eating (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**% Rear**
Preference	1	65	1	66	2
	0	128	4	132	3
	2	33	0	33	0
Treatment	C	66	0	66	0
	G	64	2	66	3
	P	63	3	66	5
	S	33	0	33	0
Replicate	1	75	2	77	3
	2	75	2	77	3
	3	76	1	77	1
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	21	0	21	0
	Phoenix	18	3	21	14
	Pixie Fae	20	1	21	5
	Semitone	20	1	21	5
	Shane	21	0	21	0
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**226**	**5**	**231**	**2**

A.6.5. Chin

Chin wobble

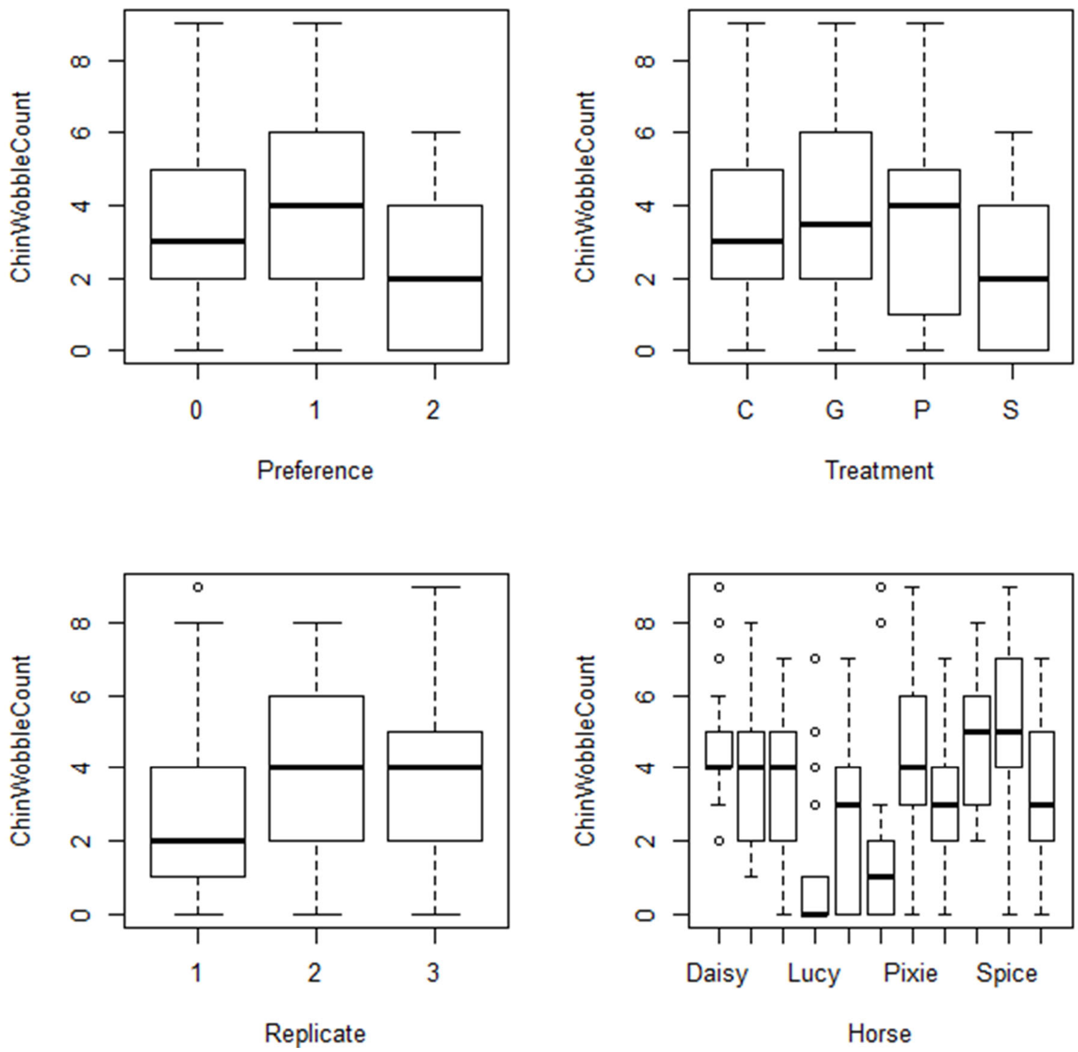


Figure A 41 Boxplots showing the chin wobble count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Parrot mouth

Table A 34 Two-way table showing the number of times a parrot mouth did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Parrot mouth (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	128	4	132	3
	2	32	1	33	3
Treatment	C	64	2	66	3
	G	64	2	66	3
	P	64	2	66	3
	S	32	1	33	3
Replicate	1	74	3	77	4
	2	74	3	77	4
	3	76	1	77	1
Horse	Daisy	20	1	21	5
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	21	0	21	0
	Phoenix	17	4	21	19
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	20	1	21	5
	Whitney	20	1	21	5
Total		**224**	**7**	**231**	**3**

Puckered chin

**Table A 35 Two-way table showing the number of times a puckered chin did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.**

		Puckered chin (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	64	2	66	3
	0	123	9	132	7
	2	32	1	33	3
Treatment	C	64	2	66	3
	G	60	6	66	9
	P	63	3	66	5
	S	32	1	33	3
Replicate	1	74	3	77	4
	2	72	5	77	6
	3	73	4	77	5
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	19	2	21	10
	Minty	21	0	21	0
	Phoenix	17	4	21	19
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	20	1	21	5
	Spice	20	1	21	5
	Whitney	17	4	21	19
Total		**219**	**12**	**231**	**5**

A.6.6. Cheek

Cheek muscle contraction

Table A 36 Two-way table showing the number of times a cheek muscle contraction did and did not occur by preference, treatment, replicate, and horse. The % column shows the percentage of observations for which the behaviour was observed.

		Cheek muscle contraction (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	65	1	66	2
	0	132	0	132	0
	2	30	3	33	9
Treatment	C	66	0	66	0
	G	66	0	66	0
	P	65	1	66	2
	S	30	3	33	9
Replicate	1	76	1	77	1
	2	76	1	77	1
	3	75	2	77	3
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	21	0	21	0
	Minty	20	1	21	5
	Phoenix	18	3	21	14
	Pixie Fae	21	0	21	0
	Semitone	21	0	21	0
	Shane	21	0	21	0
	Spice	21	0	21	0
	Whitney	21	0	21	0
Total		**227**	**4**	**231**	**2**

A.6.7. Nares

Nares flared

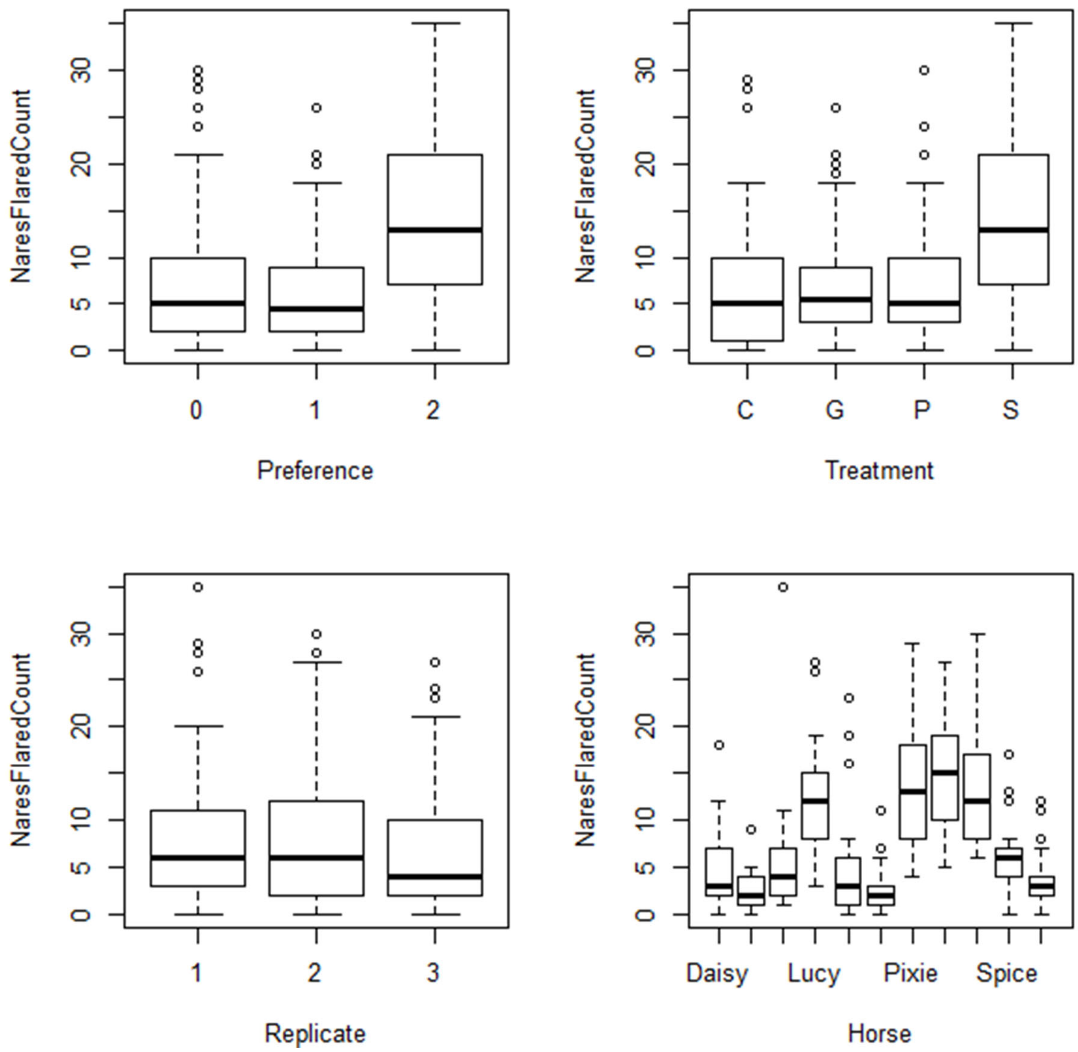


Figure A 42 Boxplots showing the nares flared count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

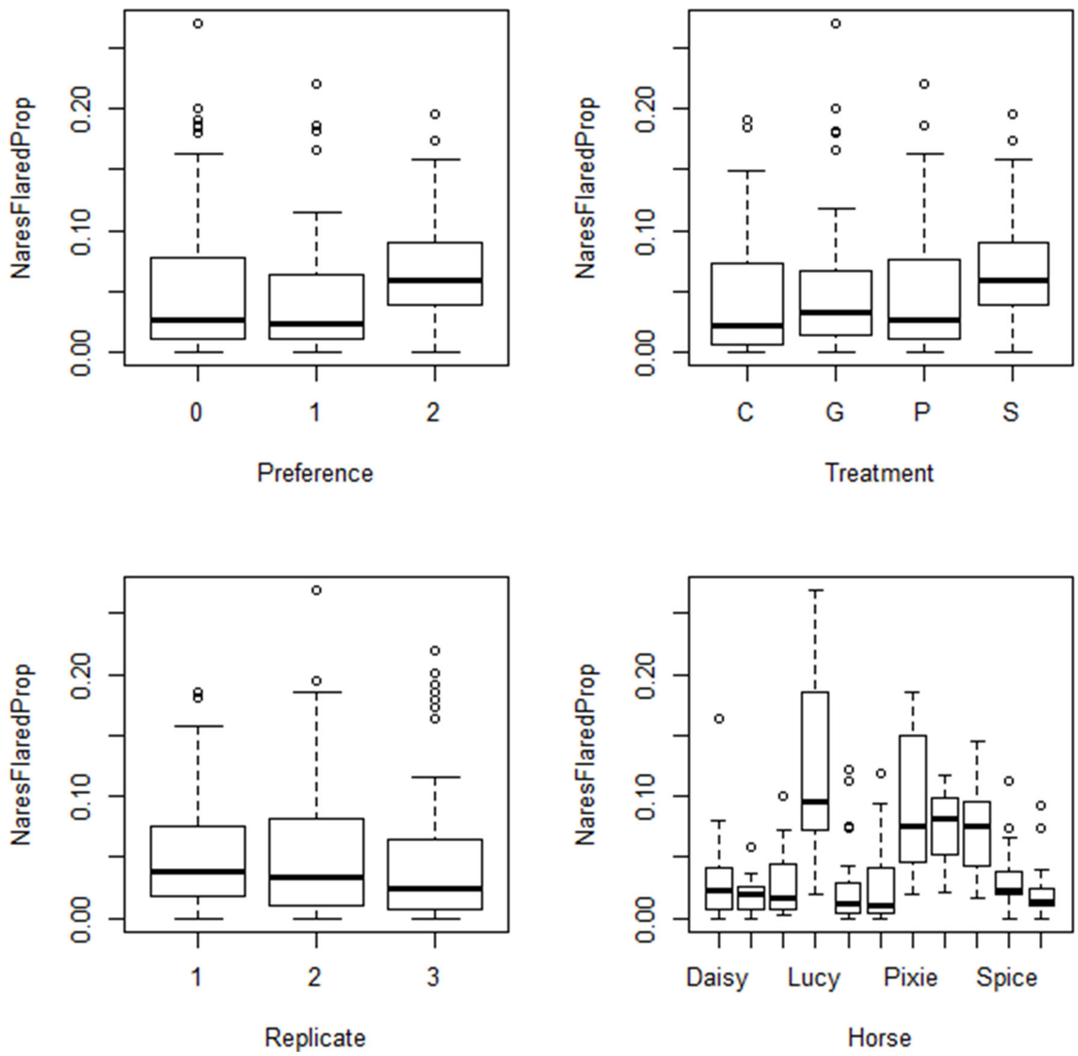


Figure A 43 Boxplots showing the nares flared proportion of time data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Nares neutral

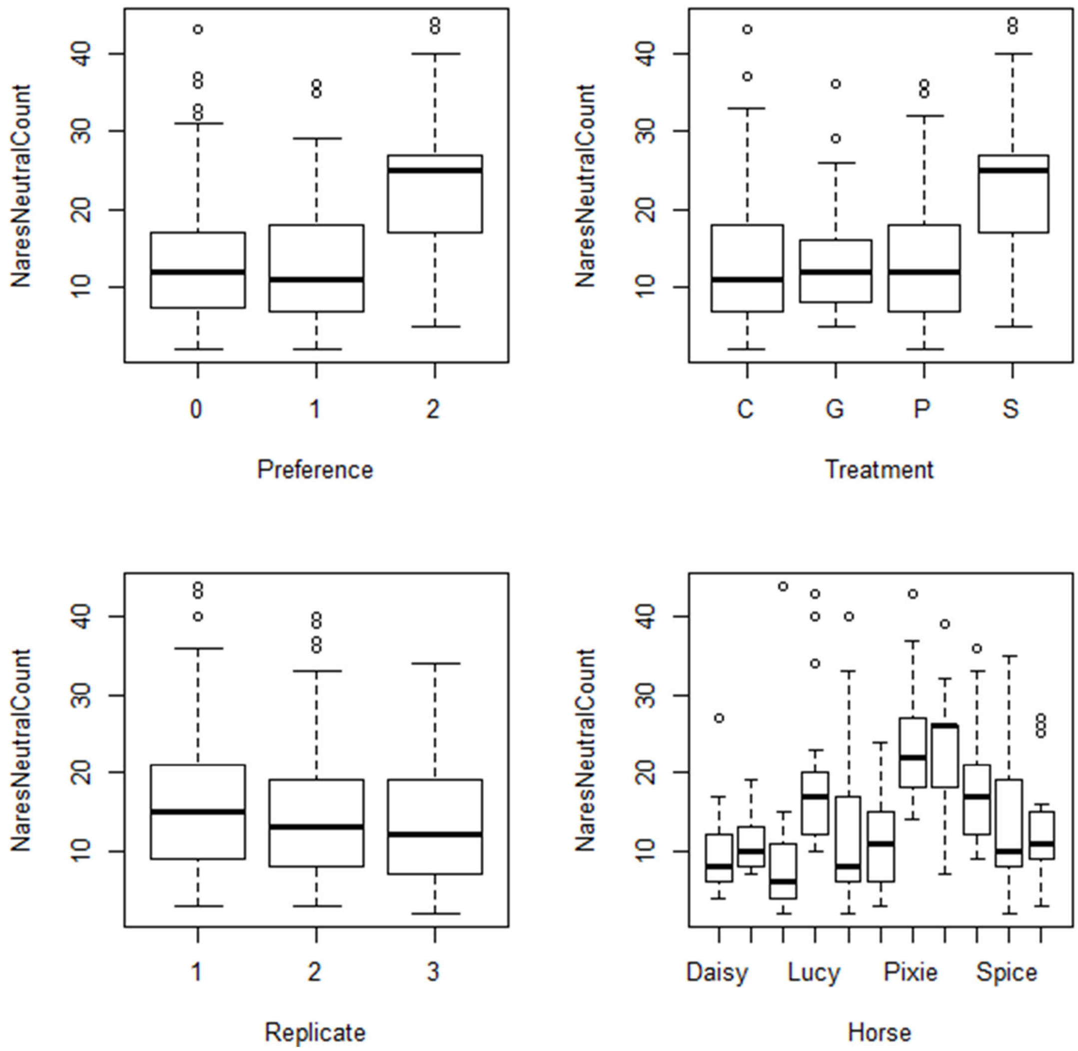


Figure A 44 Boxplots showing the nares neutral count data stratified by (clockwise from top left) preference, treatment, horse, and replicate.
The observed median, first and third quartiles, range of data, and potential outliers (circles) are represented. Preference: 0 = Intermediate, 1 = Preferred, 2 = Aversive. Treatment: C = No-stimulus, G = Groom, P= Person, S= Spray.

Nares narrowed

Table A 37 Two-way table showing the number of times a nares narrowed did and did not occur by preference, treatment, replicate, and horse.
The % column shows the percentage of observations for which the behaviour was observed.

		Nares narrowed (n)			
Variable	**Level**	**No**	**Yes**	**Total**	**%**
Preference	1	55	11	66	17
	0	112	20	132	15
	2	26	7	33	21
Treatment	C	57	9	66	14
	G	54	12	66	18
	P	56	10	66	15
	S	26	7	33	21
Replicate	1	63	14	77	18
	2	64	13	77	17
	3	66	11	77	14
Horse	Daisy	21	0	21	0
	Frankie	21	0	21	0
	Grace	21	0	21	0
	Lucy	10	11	21	52
	Minty	19	2	21	10
	Phoenix	6	15	21	71
	Pixie Fae	16	5	21	24
	Semitone	20	1	21	5
	Shane	20	1	21	5
	Spice	21	0	21	0
	Whitney	18	3	21	14
Total		**193**	**38**	**231**	**16**

A.6.1. Analysis by preference

Table A 38 Results from models examining the effect of preference for a treatment on the occurrence, frequency, or proportion of time spent displaying various head behaviours.
Analysis is based on comparison to the Preferred treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Head	Shake [b^]	Preferred	REF				19.9%
		Intermediate	0.48 (0.37)	-	1.61 (0.79-3.30)	0.192	28.6%
		Aversive	0.44 (0.50)	-	1.55 (0.58-4.15)	0.385	27.8%
	Tossing [b^]	Preferred	REF				8.2%
		Intermediate	0.22 (0.48)	-	1.25 (0.49-3.18)	0.641	10.0%
		Aversive	1.10 (0.61)	-	2.99 (0.91-9.80)	0.070	21.1%
Ears	Left forward count [a]	Preferred	REF				12.7
		Intermediate	***0.13 (0.04)***	***1.13 (1.05-1.22)***	***-***	***<0.001****	***14.4***
		Aversive	***0.20 (0.05)***	***1.23 (1.11-1.36)*** [e]	***-***	***<0.001****	***15.6***
	Right forward count [a]	Preferred	REF				14.2
		Intermediate	**0.09 (0.04)**	**1.10 (1.02-1.18)**	**-**	**0.012***	**15.6**
		Aversive	***-0.15 (0.06)***	***0.86 (0.77-0.97)***	***-***	***0.010****	***12.3***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Left forward time [c]	Intercept (Preferred)	0.25 (0.05)	-	-	<0.001	25.4%
		Intermediate	0.03 (0.02)	-	-	0.129	28.1%
		Aversive	**-0.06 (0.03)** [f]	**-**	**-**	**0.017***	**19.8%**
	Right forward time [c]	Intercept (Preferred)	0.23 (0.04)	-	-	<0.001	23.0%
		Intermediate	0.02 (0.02)	-	-	0.330	24.7%
		Aversive	***-0.11 (0.03)***	**-**	**-**	***<0.001****	***11.8%***
	Left side count [a]	Preferred	REF				17.9
		Intermediate	0.01 (0.03)	1.01 (0.94-1.08)	-	0.828	18.1
		Aversive	**-0.10 (0.05)**	**0.91 (0.82-1.00)**	**-**	**0.057***	**16.3**
	Right side count [a]	Preferred	REF				17.7
		Intermediate	**0.07 (0.03)**	**1.07 (1.00-1.14)**	**-**	**0.058***	**18.9**
		Aversive	**-0.12 (0.05)**	**0.88 (0.80-0.98)**	**-**	**0.022***	**15.6**
	Left side time [c]	Intercept (Preferred)	0.37 (0.04)	-	-	<0.001	36.7%
		Intermediate	-0.01 (0.02)	-	-	0.788	36.1%
		Aversive	***-0.18 (0.03)***	**-**	**-**	***<0.001****	***18.6%***
	Right side time [c]	Intercept (Preferred)	0.37 (0.04)	-	-	<0.001	36.6%
		Intermediate	0.00 (0.02)	-	-	0.938	36.8%

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
		Aversive	***-0.16 (0.03)***	***-***	***-***	***<0.001****	***20.4%***
	Left back count [a]	Preferred	REF				9.9
		Intermediate	**0.11 (0.04)**	**1.12 (1.02-1.22)**	**-**	**0.012***	**11.0**
		Aversive	***0.58 (0.05)***	***1.78 (1.60-1.98)***	***-***	***<0.001****	***17.6***
	Right back count [a]	Preferred	REF				10.5
		Intermediate	***0.19 (0.04)***	***1.21 (1.11-1.31)***	***-***	***<0.001****	***12.7***
		Aversive	***0.54 (0.05)***	***1.72 (1.55-1.91)***	***-***	***<0.001****	***18.1***
	Left back time [c]	Intercept (Preferred)	0.35 (0.07)	-	-	<0.001	35.4%
		Intermediate	-0.01 (0.03)	-	-	0.715	34.4%
		Aversive	***0.26 (0.04)***	***-***	***-***	***<0.001****	***61.1%***
	Right back time [c]	Intercept - Preferred	0.38 (0.06)	-	-	<0.001	38.2%
		Intermediate	-0.01 (0.03)	-	-	0.793	37.5%
		Aversive	***0.29 (0.04)***	***-***	***-***	***<0.001****	***67.2%***
	Left flat [b^]	Preferred	REF				2.0%
		Intermediate	0.00 (0.71)	-	1.00 (0.25-4.05)	1.000	2.0%
		Aversive	1.02 (0.87)	-	2.77 (0.50-15.18)	0.241	5.4%

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right flat [b^]	Preferred	REF				2.4%
		Intermediate	0.00 (0.70)	-	1.00 (0.25-3.97)	1.000	2.4%
		Aversive	1.33 (0.82)	-	3.78 (0.75-19.00)	0.106	8.6%
Eye	Blink count [a]	Preferred	REF				41.1
		Intermediate	***0.06 (0.02)***	***1.07 (1.02-1.12)***	***-***	***0.008****	***43.8***
		Aversive	***0.16 (0.03)***	***1.17 (1.10-1.25)***	***-***	***<0.001****	***48.1***
	Sclera count [a]	Preferred	REF				3.9
		Intermediate	-0.08 (0.07)	0.92 (0.81-1.05)	-	0.221	3.6
		Aversive	***0.26 (0.08)***	***1.30 (1.10-1.53)***	***-***	***0.002****	***5.1***
	Aperture small count [a]	Preferred	REF				0.9
		Intermediate	***-0.58 (0.14)***	***0.56 (0.42-0.74)***	***-***	***<0.001***	***0.5***
		Aversive	***-0.85 (0.25)***	***0.43 (0.26-0.70)***	***-***	***<0.001***	***0.4***
	Aperture neutral count [a]	Preferred	REF				6.8
		Intermediate	-0.09 (0.06)	0.91 (0.81-1.02)	-	0.119	6.2
		Aversive	***0.35 (0.07)***	***1.42 (1.23-1.64)***	***-***	***<0.001****	***9.6***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Aperture large [b^]	Preferred	REF				1.7%
		Intermediate	-0.33 (0.70)	-	0.72 (0.18-2.82)	0.635	1.2%
		Aversive	***3.33 (0.90)***	***-***	***27.97 (4.83-162.15)***	***<0.001****	***32.6%***
	Eyebrow wrinkle count [a]	Preferred	REF				1.7
		Intermediate	0.09 (0.11)	1.09 (0.89-1.34)	-	0.402	1.8
		Aversive	**0.27 (0.14)**	**1.31 (1.00-1.72)**	**-**	**0.050***	**2.2**
	Eyebrow angle count [a]	Preferred	REF				1.4
		Intermediate	**0.24 (0.12)**	**1.27 (1.01-1.61)**	**-**	**0.039***	**1.8**
		Aversive	***0.64 (0.14)***	***1.89 (1.43-2.50)***	***-***	***<0.001****	***2.7***
Mouth	Contracted lips [b^]	Preferred	REF				8.3%
		Intermediate	**-1.14 (0.59)**	**-**	**0.32 (0.10-1.00)**	**0.051***	**2.8%**
		Aversive	***2.46 (0.61)***	***-***	***11.74 (3.57-38.55)***	***<0.001****	***51.5%***
	Drooping lip [b^]	Preferred	REF				33.1%
		Intermediate	-0.06 (0.44)	-	0.94 (0.40-2.21)	0.884	31.8%
		Aversive	***-1.74 (0.68)***	***-***	***0.18 (0.05-0.66)***	***0.010****	***8.0%***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Lip licking & chewing count [a]	Preferred	REF				4.0
		Intermediate	-0.03 (0.07)	0.97 (0.84-1.12)	-	0.681	3.9
		Aversive	***-0.65 (0.13)***	***0.52 (0.41-0.67)***	***-***	***<0.001****	***2.1***
	Face pulling [b^]	Preferred	REF				5.0%
		Intermediate	-0.71 (0.65)	-	0.49 (0.14-1.77)	0.277	2.5%
		Aversive	-1.68 (1.13)	-	0.19 (0.02-1.71)	0.138	1.0%
	Oral investigation count [a]	Preferred	REF				1.2
		Intermediate	***-0.35 (0.11)***	***0.71 (0.57-0.87)***	***-***	***0.001****	***0.8***
		Aversive	***-1.45 (0.25)***	***0.23 (0.14-0.38)***	***-***	***<0.001****	***0.3***
Chin	Chin wobble count [a]	Preferred	REF				4.0
		Intermediate	**-0.16 (0.08)**	**0.85 (0.73-0.99)**	**-**	**0.035***	**3.4**
		Aversive	***-0.54 (0.13)***	***0.58 (0.45-0.74)***	***-***	***<0.001****	***2.3***
	Puckered chin [b^]	Preferred	REF				0.0%
		Intermediate	1.42 (1.03)	-	4.12 (0.55-30.93)	0.169	0.1%
		Aversive	0.00 (1.51)	-	1.00 (0.05-19.21)	1.000	0.0%

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Nares	Flared count [a]	Preferred	REF				6.3
		Intermediate	**0.11 (0.06)**	**1.12 (1.00-1.25)**	**-**	**0.053***	**7.0**
		Aversive	***0.82 (0.07)***	***2.27 (2.00-2.59)***	***-***	***<0.001****	***14.2***
	Flared time [c]	Intercept (Preferred)	0.04 (0.01)	-	-	<0.001	4.3%
		Intermediate	0.01 (0.01)	-	-	0.257	4.9%
		Aversive	***0.03 (0.01)***	***-***	***-***	***<0.001****	***7.0%***
	Neutral count [a]	Preferred	REF				14.0
		Intermediate	0.08 (0.04)	1.08 (1.00-1.17)	-	0.062	15.1
		Aversive	***0.68 (0.05)***	***1.98 (1.79-2.18)***	***-***	***<0.001****	***27.6***
	Narrowed [b^]	Preferred	REF				6.0%
		Intermediate	-0.18 (0.51)	-	0.84 (0.31-2.29)	0.731	5.1%
		Aversive	0.48 (0.69)	**-**	1.61 (0.42-6.19)	0.487	**9.4%**

* $p \leq 0.05$
^ Caution with interpretation as large SE
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the Preferred treatment, the odds of a head shake occurring was 0.45 (95%CI 0.19-1.03) times lower during the Aversive treatment and this was not statistically significant (p = 0.059). There was an 11.0% predicted probability of occurrence of a head shake during the Aversive treatment.
[e] Compared to the Preferred treatment, the rate of left ear forward behaviour was 1.23 (95%CI 1.11 - 1.36) times greater during the Aversive treatment and

this was statistically significant ($p < 0.001$). Left ear forward was predicted to occur 15.6 times during the three-minute Aversive treatment.
[f] Compared to the Preferred treatment, the proportion of time that horses had their left ear forward was 0.06 (SE 0.03) less during the Aversive treatment, this was statistically significant ($p = 0.017$). The predicted percentage of time that horses had their left ear forward was 25.4% for the Preferred treatment and 19.4% for the Aversive treatment.

A.6.2. Analysis by treatment

Table A 39 Results from models examining the effect of treatment on the occurrence, frequency, or proportion of time spent displaying various head behaviours.
Analysis is based on comparison to the No-stimulus treatment (referent). Where p ≤ 0.05 results are shown in bold, where p ≤ 0.01 results are also italicised.

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Head	Shake [b^]	No-stimulus	REF				18.4%
		Groom	***1.08 (0.42)***	***-***	***2.94 (1.29-6.68)***	***0.010****	***39.4%***
		Person	0.10 (0.44)	-	1.10 (0.47-2.61)	0.822	19.6%
		Spray	0.54 (0.51)	-	1.72 (0.63-4.67)	0.289	27.5%
	Tossing [b^]	No-stimulus	REF				13.2%
		Groom	-0.72 (0.55)	-	0.49 (0.17-1.43)	0.190	6.9%
		Person	-0.55 (0.53)	-	0.58 (0.20-1.63)	0.300	8.1%
		Spray	0.55 (0.57)	-	1.74 (0.57-5.33)	0.334	20.9%
Ears	Left forward count [a]	No-stimulus	REF				14.7
		Groom	-0.08 (0.04)	0.93 (0.85-1.01)	-	0.074	13.7
		Person	***-0.12 (0.04)***	***0.89 (0.81-0.96)*** [e]	***-***	***0.005****	***13.1***
		Spray	0.05 (0.05)	1.06 (0.96-1.16)	-	0.271	15.6

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right forward count [a]	No-stimulus	REF				15.7
		Groom	0.03 (0.04)	1.03 (0.95-1.11)	-	0.488	16.2
		Person	***-0.15 (0.04)***	***0.86 (0.79-0.94)***	***-***	***<0.001****	***13.5***
		Spray	***-0.25 (0.06)***	***0.78 (0.70-0.87)***	***-***	***<0.001****	***12.3***
	Left forward time [c]	Intercept (No-stimulus)	0.28 (0.05)	-	-	<0.001	27.9%
		Groom	0.02 (0.02)	-	-	0.439	29.5%
		Person	-0.04 (0.02)	-	-	0.071	24.2%
		Spray	***-0.09 (0.02)*** [f]	***-***	***-***	***<0.001****	***19.4%***
	Right forward time [c]	Intercept (No-stimulus)	0.26 (0.04)	-	-	<0.001	26.2%
		Groom	0.00 (0.02)	-	-	0.833	26.6%
		Person	***-0.07 (0.02)***	***-***	***-***	***0.001****	***19.6%***
		Spray	***-0.14 (0.02)***	***-***	***-***	***<0.001****	***11.8%***
	Left side count [a]	No-stimulus	REF				18.0
		Groom	0.00 (0.04)	1.00 (0.92-1.08)	-	0.961	18.0
		Person	0.00 (0.04)	1.00 (0.93-1.08)	-	0.965	18.1
		Spray	**-0.10 (0.05)**	**0.90 (0.82-1.00)**	**-**	**0.044***	**16.3**

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right side count [a]	No-stimulus	REF				19.1
		Groom	-0.02 (0.04)	0.98 (0.91-1.06)	-	0.664	18.8
		Person	***-0.08 (0.04)***	***0.92 (0.85-1.00)***	***-***	***0.040****	***17.6***
		Spray	***-0.20 (0.05)***	***0.82 (0.74-0.91)***	***-***	***<0.001****	***15.6***
	Left side time [c]	Intercept (No-stimulus)	0.36 (0.04)	-	-	<0.001	35.8%
		Groom	0.03 (0.02)	-	-	0.255	38.5%
		Person	-0.01 (0.02)	-	-	0.664	34.7%
		Spray	***-0.17 (0.03)***	***-***	***-***	***<0.001****	***18.6%***
	Right side time [c]	Intercept (No-stimulus)	0.37 (0.04)	-	-	<0.001	36.6%
		Groom	0.01 (0.02)	-	-	0.695	37.5%
		Person	0.00 (0.02)	-	-	0.880	36.2%
		Spray	***-0.16 (0.03)***	***-***	***-***	***<0.001****	***20.4%***
	Left back count [a]	No-stimulus	REF				10.7
		Groom	**-0.10 (0.05)**	**0.91 (0.82-1.00)**	**-**	**0.053***	**9.7**
		Person	0.08 (0.05)	1.08 (0.98-1.19)	-	0.121	11.5
		Spray	***0.50 (0.05)***	***1.64 (1.48-1.82)***	***-***	***<0.001****	***17.6***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right back count [a]	No-stimulus	REF				12.3
		Groom	-0.09 (0.05)	0.92 (0.83-1.01)	-	0.068	11.3
		Person	0.00 (0.05)	1.00 (0.91-1.09)	-	0.957	12.3
		Spray	***0.38 (0.05)***	***1.47 (1.32-1.63)***	***-***	***<0.001****	***18.1***
	Left back time [c]	Intercept (No-stimulus)	0.36 (0.07)	-	-	<0.001	35.7%
		Groom	**-0.06 (0.03)**	**-**	**-**	**0.050***	**29.8%**
		Person	0.03 (0.03)	-	-	0.332	38.6%
		Spray	***0.25 (0.04)***	***-***	***-***	***<0.001****	***61.1%***
	Right back time [c]	Intercept (No-stimulus)	0.37 (0.06)	-	-	<0.001	37.1%
		Groom	-0.03 (0.03)	-	-	0.322	34.0%
		Person	0.05 (0.03)	-	-	0.108	42.1%
		Spray	***0.30 (0.04)***	***-***	***-***	***<0.001****	***67.2%***
	Left flat [b^]	No-stimulus	REF				0.3%
		Groom	2.06 (1.22)	-	7.81 (0.72-84.72)	0.091	2.4%
		Person	**2.34 (1.21)**	**-**	**10.33 (0.97-110.46)**	**0.053***	**3.1%**
		Spray	**2.80 (1.29)**	**-**	**16.43 (1.32-205.10)**	**0.030***	**4.9%**

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Right flat [b^]	No-stimulus	REF				0.4%
		Groom	2.01 (1.20)	-	7.44 (0.71-78.49)	0.095	2.9%
		Person	**2.27 (1.19)**	**-**	**9.72 (0.94-100.91)**	**0.057***	**3.8%**
		Spray	**2.80 (1.26)**	**-**	**21.78 (1.86-255.38)**	**0.014***	**8.1%**
Eye	Blink count [a]	No-stimulus	REF				42.1
		Groom	***0.10 (0.03)***	***1.11 (1.05-1.17)***	***-***	***<0.001****	***46.5***
		Person	-0.05 (0.03)	0.95 (0.90-1.01)	-	0.076	40.0
		Spray	***0.13 (0.03)***	***1.14 (1.07-1.22)***	***-***	***<0.001****	***48.1***
	Sclera count [a]	No-stimulus	REF				3.8
		Groom	-0.04 (0.08)	0.96 (0.83-1.12)	-	0.610	3.7
		Person	-0.06 (0.08)	0.94 (0.81-1.10)	-	0.438	3.6
		Spray	***0.28 (0.09)***	***1.32 (1.12-1.56)***	***-***	***<0.001****	***5.1***
	Aperture small count [a]	No-stimulus	REF				0.6
		Groom	***-0.54 (0.20)***	***0.58 (0.40-0.86)***	***-***	***0.007****	***0.4***
		Person	**0.39 (0.16)**	**1.48 (1.08-2.03)**	**-**	**0.014***	**0.9**
		Spray	-0.49 (0.26)	0.61 (0.37-1.02)	-	0.061	0.4

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Aperture neutral count [a]	No-stimulus					6.1
		Groom	0.02 (0.07)	1.02 (0.89-1.17)	-	0.795	6.3
		Person	0.09 (0.07)	1.10 (0.96-1.26)	-	0.175	6.7
		Spray	***0.45 (0.07)***	***1.57 (1.35-1.81)***	***-***	***<0.001****	***9.6***
	Aperture large [b^]	No-stimulus	REF				0.8%
		Groom	0.41 (0.91)	-	1.50 (0.25-8.88)	0.654	1.2%
		Person	1.03 (0.86)	-	2.80 (0.52-15.26)	0.233	2.1%
		Spray	***4.12 (1.04)***	***-***	***61.49 (7.98-474.01)***	***<0.001****	***32.4%***
	Eyebrow wrinkle count [a]	No-stimulus	REF				1.8
		Groom	-0.04 (0.12)	0.96 (0.76-1.21)	-	0.741	1.8
		Person	-0.11 (0.12)	0.90 (0.71-1.14)	-	0.383	1.7
		Spray	0.17 (0.14)	1.18 (0.90-1.54)	-	0.226	2.2
	Eyebrow angle count [a]	No-stimulus	REF				1.4
		Groom	***0.43 (0.13)***	***1.53 (1.19-1.97)***	***-***	***<0.001****	***2.1***
		Person	0.08 (0.14)	1.09 (0.82-1.43)	-	0.560	1.5
		Spray	***0.66 (0.14)***	***1.93 (1.46-2.56)***	***-***	***<0.001****	***2.7***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
Mouth	Contracted lips [b^]	No-stimulus	REF				2.8%
		Groom	-0.44 (0.95)	-	0.65 (0.10-4.14)	0.6644	1.8%
		Person	1.30 (0.72)	-	3.67 (0.90-14.96)	0.070	9.4%
		Spray	***3.62 (0.78)***	**-**	***37.43 (8.17-171.53)***	***<0.001****	***51.5%***
	Drooping lip [b^]	No-stimulus	REF				32.7%
		Groom	0.39 (0.51)	-	1.48 (0.54-4.03)	0.446	41.8%
		Person	-0.52 (0.51)	-	0.59 (0.22-1.63)	0.312	22.4%
		Spray	***-1.77 (0.69)***	**-**	***0.17 (0.04-0.65)***	***0.010****	***7.6%***
	Lip licking & chewing count [a]	No-stimulus	REF				3.8
		Groom	**0.19 (0.08)**	**1.20 (1.03-1.41)**	**-**	**0.023***	**4.6**
		Person	-0.14 (0.09)	0.87 (0.73-1.04)	-	0.116	3.3
		Spray	***-0.61 (0.13)***	***0.55 (0.42-0.70)***	**-**	***<0.001****	***2.1***
	Face pulling [b^]	No-stimulus	REF				2.7%
		Groom	0.82 (0.76)	-	2.27 (0.51-10.05)	0.279	5.9%
		Person	-0.73 (0.87)	-	0.48 (0.09-2.66)	0.402	1.3%
		Spray	-1.18 (1.16)	**-**	0.31 (0.03-3.01)	0.311	0.8%

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Oral investigation count [a]	No-stimulus	REF				0.5
		Groom	***0.71 (0.15)***	***2.03 (1.50-2.74)***	***-***	***<0.001****	***1.0***
		Person	***0.95 (0.15)***	***2.60 (1.94-3.48)***	***-***	***<0.001****	***1.3***
		Spray	**-0.60 (0.27)**	**0.55 (0.32-0.93)**	**-**	**0.026***	**0.3**
Chin	Chin wobble count [a]	No-stimulus	REF				3.6
		Groom	0.03 (0.09)	1.03 (0.86-1.23)	-	0.746	3.7
		Person	-0.07 (0.09)	0.94 (0.78-1.12)	-	0.471	3.4
		Spray	***-0.45 (0.13)***	***0.64 (0.50-0.82)***	***-***	***<0.001****	***2.3***
	Puckered chin [b^]	No-stimulus	REF				0.0
		Groom	2.12 (1.17)	-	8.31 (0.83-83.02)	0.071	0.1
		Person	0.69 (1.19)	-	1.99 (0.19-20.37)	0.564	0.0
		Spray	0.00 (1.54)	-	1.00 (0.05-20.26)	1.000	0.0
Nares	Flared count [a]	No-stimulus	REF				6.4
		Groom	-0.01 (0.07)	0.99 (0.87-1.13)	-	0.927	6.4
		Person	**0.14 (0.07)**	**1.15 (1.01-1.31)**	**-**	**0.036***	**7.4**
		Spray	***0.79 (0.07)***	***2.20 (1.94-2.51)***	***-***	***<0.001****	***14.2***

Head part	Behaviour	Variable	Beta (SE)	Rate ratio (95%CI)	Odds ratio (95%CI)	p-value (Wald)	Predicted value[#]
	Flared time [c]	Intercept (No-stimulus)	0.04 (0.01)	-	-	<0.001	4.2%
		Groom	0.01 (0.01)	-	-	0.201	5.0%
		Person	0.01 (0.01)	-	-	0.204	5.0%
		Spray	***0.03 (0.01)***	***-***	***-***	***<0.001****	***7.0%***
	Neutral count [a]	No-stimulus	REF				14.6
		Groom	-0.06 (0.05)	0.95 (0.86-1.04)	-	0.251	13.8
		Person	0.09 (0.05)	1.09 (0.99-1.20)	-	0.072	15.9
		Spray	***0.64 (0.05)***	***1.89 (1.72-2.09)***	***-***	***<0.001****	***27.6***
	Narrowed [b^]	No-stimulus	REF				4.6%
		Groom	0.53 (0.60)	-	1.70 (0.52-5.51)	0.377	7.0%
		Person	0.19 (0.61)	-	1.21 (0.36-4.00)	0.760	5.0%
		Spray	0.84 (0.71)	-	2.32 (0.58-9.32)	0.234	9.3%

* $p \leq 0.05$
^ Caution with interpretation as SE may be large
[#] Model predicted value for binary data is probability of the behaviour occurring, for count data it is the predicted count for a three-minute treatment period, and for behaviour duration data it is the predicted proportion of time.
[a] Count data analysed using a mixed effects Poisson model.
[b] Binary data analysed using a mixed effects logistic regression model.
[c] Proportion of time data analysed using a mixed linear regression model.
[d] Compared to the No-stimulus treatment, the odds of a head shake occurring was 2.90 (95%CI 1.31 - 6.42) times higher during the Groom treatment and this was statistically significant (p = 0.009). There was a 39.5% predicted probability of occurrence of a head shake during the Groom treatment.
[e] Compared to the No-stimulus treatment, the rate of left ear forward behaviour was 0.89 (95%CI 0.81 - 0.96) times less during the Person treatment, this was

statistically significant ($p = 0.005$). Left ear forward was predicted to occur 13.1 times during the three-minute Person treatment.
[f] Compared to the No-stimulus treatment, the proportion of time that horses had their left ear forward was 0.09 (SE 0.02) less during the Spray treatment, this was statistically significant ($p < 0.001$). The predicted percentage of time that horses had their left ear forward was 27.9% for the No-stimulus treatment and 19.4% for the Spray treatment.

A.7. Chapter 11 Supplementary material

Table A 40 Results from exploratory factor analysis using PCF (no rotation) of the 16 parameters that varied across the preference levels during univariable. Loading values between -0.3900 and 0.3900 were removed. Factors 1- 5 collective explain 67.3% of the variation in the data.

Parameter	Factor 1 26.8%	Factor 2 12.4%	Factor 3 12.0%	Factor 4 8.8%	Factor 5 7.3%
Nares neutral	0.7301				
Nares flared	0.7288				
Left ear forward	0.6930	-0.5050			
Right ear forward	0.6310	-0.6673			
Right ear back	0.6256		-0.5266		
Left ear back	0.6227		-0.5589		
Neck very low	0.5897		0.5407		
Right ear side	0.5285	-0.5275	0.4404		
Eyebrow angled	0.5119				
Neck very high	0.4499			-0.6197	
Tail swish	0.4169	0.4979		-0.4077	
Small eye aperture	-0.3907				0.5409
Mouth contracted		0.4942	0.3256	0.4152	
Chin wobble		-0.4237	-0.3966		
Oral investigation			0.5761		0.4959
Blink				0.5260	-0.5471

Table A 41 Results of principle component analysis (PCA) of the 16 parameters that varied across the preference levels during univariable analysis.
Five components were retained in the model. One parameter (blink) was removed as it did not meet the threshold for inclusion (i.e., factor loading was between -0.2000 and 0.2000) in any of the factors. Principal components (PC) 1-5 collective explain 70.5% of the variation in the data.

Parameter	PC1 28.5% Nares-ears	PC2 13.2% Ear forward-side	PC3 12.8% Oral-neck low-ear back	PC4 8.9% Neck high-oral-tail	PC5 7.1% Eye small-nares
Nares flared	0.3535				-0.2631
Nares neutral	0.3531				-0.3795
Left ear forward	0.3345	-0.3619			
Right ear forward	0.3057	-0.4752			
Right ear back	0.3027		-0.3704		0.3471
Left ear back	0.3003	0.2395	-0.3924	0.2084	0.2707
Neck very low	0.2856		0.3981		0.2899
Right ear side	0.2558	-0.3889	0.2993		
Eyebrow angled	0.2477				
Neck very high	0.2192		0.2162	-0.5203	0.2583
Tail swish	0.2018	0.3593		-0.3632	
Mouth contracted		0.3243	0.2457	0.3346	-0.2822
Chin wobble		-0.2711	-0.2941	-0.2324	0.2760
Oral investigation			0.4119	0.4617	
Small eye aperture				0.3071	0.4874