AN INVESTIGATION INTO PHYSIOLOGICAL CORRELATES OF EQUINE PERSONALITY.

Aurélie JOLIVALD

A thesis submitted in partial fulfilment of the requirements of
Nottingham Trent University for the degree of Doctor of
Philosophy

September 2021

Acknowledgements

First of all I would like to thank my supervisors Dr Kelly Yarnell, Dr Carrie Ijichi and Dr Carol Hall. Of course, thank you so much for your scientific guidance and support over the past four years. But also, thank you for letting me take this project in all sorts of weird and wonderful directions. Thank you for magicking lab equipment into existence and bullying me into learning to appreciate Gantt charts. Thank you especially for your moral support towards the end. I am truly grateful for your support and for all the opportunities you helped me take.

I am very grateful to Nottingham Trent University for funding the studentship that allowed me to pursue this project.

I would also like to thank all the technical team at Brackenhurst Equestrian Centre for their welcome and their support. In particular, thank you to Anna Gregory for arranging facilities, horses and technical assistance for me over the last four years. A special thank you to Jude Lucas, who provided technical support during data collection, and to Amy Hazelhurst, Cath Hake, Catherine Rhoades and Jake Bromley-Fowles, who provided personality assessment for the horses under their care.

Thank you also to the NTU Equine lecturing team, who welcomed me with open arms from the start of my PhD. Thank you so much for trusting me with some of your sessions and allowing me to gain teaching experience. A special thank you to Sarah Hallam, Tina Canton and Liz Taylor for providing personality assessments for the horses as well.

Thank you to Dr Jill Labadz and Prof Carl Smith for their support and guidance as ARES PGR tutors. In addition, I am extremely grateful to Carl for his help with the statistics in this project.

Thank you to Jen Burton who provided help with data collection as part of her dissertation project. Thank you also to Lucile Vigouroux, the best MSc assistant ever. I am so sorry Covid

meant the screen project had to be cut out of this thesis, but I remain hugely thankful for every minute of your help, and very proud of the success you've had with it.

To everyone in the PhD office over the years, thank you for your friendship and support, for all the laughs, for the cake days, for the PhD socials, for the summer lunches by the Brack pond, and even for the days that required lying on the floor with the dog. I could not have done this PhD without you – and most importantly, even if I could have it would not have been anywhere near as much fun! A special thanks to Anthony, Ellen, Kate and Kym for their help with data collection at various points - and to Alex and Kate for their incredible online support in the very last stretch…

I also have amazing friends outside of NTU to thank. Lisa, Marion, Terri, thank you so much for your unending support and belief, and for always being but a message away, rain or shine. To everyone at Bet's yard, thank you for being my home away from home for the past 6 years. And of course thank you to Cabral, my emotional support pony all the way through university.

Finalement (mais c'est le plus important!), je voudrais remercier ma famille. Mes koudous, papi et mamie, Mamiemone et papi Pierrot, merci pour votre soutien sans faille, moral et pratique, pour tous les coups de téléphone et les voyages en Angleterre, pour les légumes du jardin et les mirabelles de Lorraine. Merci d'avoir toujours cru en moi et d'avoir tout fait pour m'aider, depuis toujours. Merci d'avoir fait de moi une petite fille de la campagne et d'avoir toujours soutenu mon amour des chevaux. Merci de m'avoir appris à rêver, à voyager, à persévérer, et à aimer la science. Rien de tout ça ne serait possible sans vous, et je vous dédie cette thèse.

Abstract

Objective equine personality tests enable the selection of horses for roles based on their typical behavioural responses to challenges. In humans and rodents, non-behavioural correlates of personality such as physiological reactivity to stressors and cognitive style have been identified. These traits are relevant to equine welfare and performance, yet little is known about their relationship with equine personality. Therefore, it is currently unclear what impact selection for personality has on these factors. This thesis aimed to address this gap by investigating potential neurophysiological correlates of equine personality. First, the Equine Personality Test (EPT) was evaluated for internal consistency, inter-rater reliability and test-retest reliability. This demonstrated that the EPT produces valid and reliable evaluations of the equine personality factors Agreeableness, Neuroticism, Extraversion and Gregariousness towards People. Following this, autonomic and HPA axis reactivity to stressors, chronic HPA axis activity and tonic striatal dopamine were investigated as potential neurophysiological correlates of equine personality measured by the EPT. They were measured through cardiac and salivary cortisol responses to experimental stressors, hair cortisol concentration and spontaneous blink rate, respectively. Although these physiological parameters have been established as correlates of personality in human and rodent models, no similar associations were identified in the horse. The EPT did not have predictive validity for physiological reactivity to stressors, suggesting that horses identified as non-reactive to stressors on a behavioural basis did not have equally low physiological stress sensitivity. Hair cortisol concentration was positively associated with Agreeableness, suggesting that compliant horses may experience greater HPA axis activity. These results raise concerns on the welfare of compliant, non-behaviourally reactive horses. Should they be confirmed by future research, these findings should inform the choice of methods used to select horses for roles, with a view to safeguard not only human safety but also equine welfare.

Contents

List of figures

List of tables

Chapter 1 Literature review

1.1 INTRODUCTION

The importance of personality is increasingly being recognised for both performance and leisure horses, both in the context of research (König von Borstel, 2013) and within the industry (Graf, König von Borstel and Gauly, 2013). Research suggests that the majority equestrians at all levels favour a horse that is compliant with training and has limited behavioural reactivity to stressors (Wipper, 2000; Górecka-Bruzda et al., 2011; Graf, König von Borstel and Gauly, 2013; Suwała et al., 2016). Evidence suggests that this personality type might indeed contribute to performance in horses selected for challenging roles such as police horses (Pierard, McGreevy and Geers, 2017) or competition horses (Visser et al., 2003a). Accordingly, objective field tests have been developed and are now well established to support the selection of horses for these traits on the basis of behaviour (Graf, König von Borstel and Gauly, 2014; Lansade et al., 2016). This is advantageous from the human perspective as this is expected to enhance handler safety, equine performance, and the horse-human relationship (Ijichi et al., 2013; Munsters et al., 2013b; Graf, König von Borstel and Gauly, 2013).

A tentative model of equine personality structure has also been established (Morris, Gale and Howe, 2002; McGrogan, Hutchison and King, 2008; Ijichi et al., 2013), in line with the conserved structure of personality described in other species (Gosling and John, 1999). In both human and non-human animals, the personality factors identified within this conserved structure have been shown to be driven by neurophysiological differences (Depue and Collins, 1999; Ormel et al., 2013; de Boer, Buwalda and Koolhaas, 2017). These give rise to differences in cognitive style (Coppens, de Boer and Koolhaas, 2010), physiological stress sensitivity (Koolhaas et al., 2010) or susceptibility to stress related diseases (Koolhaas, 2008) that co-vary with individual differences in behaviour. Although these covariates are highly relevant to the performance and sustainability of equines, they have not been explored in detail in the horse to date (Rankins and Wickens, 2020). In particular, little is known about the neurophysiological characteristics of different personality types in horses (Rankins and Wickens, 2020). As a result, the impact of selection based on personality on outcomes such as sensitivity to physiological stress and cognitive style is not yet well understood in the horse. This thesis aimed to start addressing this gap by exploring links between personality and potential neurophysiological correlates in the horse, with a view to establish a neurophysiological profile of the different personality types. This approach aimed to explore whether selection on the basis of personality has a positive impact on welfare, to complement its existing benefits in terms of equine performance and human safety.

1.2 FUNDAMENTAL PRINCIPLES OF PERSONALITY

1.2.1 DEFINITION

The concept of personality has originated in human psychology in order to categorise "characteristics of individuals that describe and account for consistent patterns in feeling, thinking and behaving" (Pervin and John, 1997). Personality therefore describes interrelated patterns of behaviour, affect and cognition, as well their physiological correlates (Finkemeier, Langbein and Puppe, 2018). However, subsequent research on animal models has generated ample evidence that a wide range of non-human taxa, including horses, present individual differences in behaviour and physiology that are stable over time and across contexts, consistent with the concept of personality (reviewed in e.g. Gosling, 2001; Finkemeier, Langbein and Puppe, 2018). Multiple terms are used in the animal literature to describe this phenomenon, including “behavioural syndrome” (Sih and Del Giudice, 2012), “coping style” (Koolhaas et al., 1999), or “temperament”(Réale et al., 2007). Within the context of this thesis, unless specified otherwise the terms “temperament” and “personality” will be defined in continuity with human research. Temperament is defined as “the inherited, early appearing tendencies that continue throughout life and serve as foundation to personality” (Gosling, 2001), driven by endogenous factors such as genetics and physiological processes (Scandell, 2000; Rothbart, 2007). Meanwhile, personality is defined as “A correlated set of individual behavioural and physiological traits that are consistent over time and contexts” (Finkemeier, Langbein and Puppe, 2018). Personality is generally thought to represent the result of interactions between the endogenous characteristics driving temperament, and individual experience (McCrae et al., 2000).

1.2.2 PERSONALITY MODELS IN HUMAN AND NON-HUMAN ANIMALS

Personality models can be used to describe the broad behavioural tendencies of a given individual. A characteristic stable across time and contexts is defined as a trait. Related traits are grouped into facets, which are themselves grouped into broader factors or dimensions (Gosling and John, 1999). In humans, traits may reflect behavioural, but also cognitive and affective, processes (Zillig, Hemenover and Dienstbier, 2002). Although describing the structure of personality has been the main concern of early personality science (Réale et al., 2007), a wide range of personality models are available, with little consensus over the true structure of personality (De Raad, 2009). Human personality models generally comprise three to seven broad factors (De Raad, 2009). Notable models include Cloninger’s 7 factor model (Cloninger, Svrakic and Przybeck, 1993), Eysenck’s 3 factor model (Eysenck and

Eysenck, 1975), and in particular the Five Factor Model (FFM: Mccrae and Costa, 1987). Five factor models have been extensively replicated and appear to capture a robust and comprehensive taxonomy of personality (Mccrae and Costa, 1987). Both Eysenck's and the Five Factor Model comprise the factors Neuroticism-Emotional Stability and Extraversion-Introversion (Mccrae and Costa, 1987; Eysenck and Eysenck, 1975). Neuroticism reflects an individual's sensitivity to negative affect (Mccrae and Costa, 1987) and measures facets such as anxiety, depression or vulnerability to stress (Gosling and John, 1999). By contrast, Extraversion reflects an individual's tendency for positive affect (Mccrae and Costa, 1987) and comprises facets linked with sociability but also with assertiveness, activity and positive emotions (Gosling and John, 1999). In addition, the Five Factor Model also measures another three factors. Agreeableness-Antagonism, comprising facets such as trust, tendermindedness, cooperation, and lack of aggression (Gosling and John, 1999), refers to the tendency to express pro-social attitudes and behaviour (Mccrae and Costa, 1987). Openness-Closeness to Experience, comprising facets such as intellect, imagination, creativity, and curiosity (Gosling and John, 1999), refers to an individual's curiosity, breadth of interests and daring (Mccrae and Costa, 1987). Finally, Conscientiousness vs Impulsiveness, comprising facets such as deliberation, self-discipline, dutifulness and order (Gosling and John, 1999), refers to an individual's self-control and will to achieve (Mccrae and Costa, 1987).

Individual differences in behaviour consistent with personality have been documented in a wide range of non-human animal species, ranging from non-human primates (e.g. chimpanzees: King and Figueredo, 1997; rhesus monkeys: Stevenson-Hinde and Zunz, 1978) to fish (e.g. rainbow trout: Sneddon, 2003) and invertebrates(e.g. honey bees: Walton and Toth, 2016): reviewed in Gosling (2001). Behavioural, rather than cognitive or affective traits are most often evaluated in animal models (Gosling and John, 1999). While some studies apply human personality models directly to the specie(s) of interest (e.g. Morris, Gale and Howe, 2002), the majority develop species-specific models on the basis of behavioural observations in the target species (Gosling, 2001; Kralj-Fišer and Schuett, 2014; Groothuis and Carere, 2005). Unfortunately, large variations in methodology and terminology limit the potential to generalise species-specific findings into a unified model of animal personality (Réale et al., 2007). Nonetheless, some models have been shown to be applicable in a wide range of species, highlighting their conserved character. The use of a comparative approach and a unified language across animal studies is advocated as this would facilitate cross-species and cross-discipline comparisons of findings (Gosling and John, 1999; Réale et al., 2007; Gosling, 2008).

One such highly conserved model is the coping style framework, which is concerned specifically with the different ways in which individuals react to environmental challenges (Koolhaas et al., 1999). This framework proposes that individuals vary not only in the intensity, but also in the nature of their behavioural responses to stressors. It therefore classifies individuals using a two-factor model, with orthogonal axes representing emotional reactivity and coping strategy (Koolhaas et al., 1999). The coping strategy axis describes the quality of the behavioural response to challenges and classifies individuals as proactive or reactive copers according to their tendency to respond actively, or passively, to an environmental challenge (Koolhaas et al., 1999). This framework was first developed in rodents (Henry and Stephens, 1977) but has been found to be applicable to a large range of species (Koolhaas et al., 1999), including domestic animals such as cattle or pigs (Finkemeier, Langbein and Puppe, 2018). This model has been shown to be closely linked with personality (Finkemeier, Langbein and Puppe, 2018) and provides a strong base for cross-species comparison of related traits, although it does not constitute an exhaustive model of personality traits.

Comparative studies have also suggested that some personality dimensions identified in human personality models may be conserved across a wide range of species (Gosling and

John, 1999). In reviews comparing animal personality models to the Five Factor Model, factors consistent at the trait level with some or all facets of Neuroticism, Extraversion and Agreeableness were consistently identified (Gosling and John, 1999; Gosling, 2001). In addition, factors consistent with Openness were also identified in a large proportion of species (Gosling and John, 1999), although they appear to largely represent the curiosity facet of human Openness, rather than its intellectual dimension. However, a Conscientiousness factor rarely emerges from animal studies, and was only reliably discovered in studies of chimpanzees (Gosling and John, 1999; King and Figueredo, 1997), suggesting that this dimension may be specific to humans and closely related non-human primates. The Five Factor Model therefore appears to provide a suitable base for a unified structure and language to describe animal personality (Gosling and John, 1999).

1.2.3 ASSESSMENT OF PERSONALITY

In humans, personality is most often assessed through the use of psychometric questionnaires (Boyle and Helmes, 2009). Self-report is the most common method used to obtain personality data, wherein the target individual fills in the questionnaire to describe their own patterns of affect, cognition and behaviour (Paulhus and Vazire, 2007). However, scoring by a familiar third party is also sometimes employed (McCrae et al., 2004). Examples of available psychometric instruments include Cloninger's Temperament-Character Inventory (Cloninger, Svrakic and Przybeck, 1993; Garcia et al., 2017), the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975), and the NEO Five Factor Inventory (Costa and Mccrae, 1992).

Human personality assessment tools are developed using well established guidelines for the construction of psychometric scales and must meet a number of criteria in order to be considered to produce valid and reliable measures of personality (Simms and Watson, 2007).

First, the pattern of interrelation of items making up a personality assessment tool should reflect the internal structures of personality factors, which are generally considered to be unidimensional and internally coherent constructs (Simms and Watson, 2007). All items on the scale should therefore measure the same underlying construct (termed "homogeneity") and produce sufficiently consistent scores (termed "internal consistency"). In addition, scores produced using personality assessment tools should be valid, i.e. provide an accurate reflection of the behavioural tendencies of the individual assessed and be linked with real-world behaviour or outcomes (Gosling and Vazire, 2002). Questionnaires should be evaluated for concurrent validity, i.e. how well trait ratings in the questionnaire reflect the expression of conceptually related behaviours by the target individual in real-life situations (Gosling and Vazire, 2002). They can also be evaluated for predictive validity, i.e. for how traits correlate with wider real-world outcomes (Gosling and Vazire, 2002). Finally, the personality assessment tool should satisfy several aspects of reliability. Ratings should reflect the individual's inherent behavioural tendencies rather than the rater's biases or implicit theories of personality (Gosling and Vazire, 2002). Therefore, independent raters should agree in their ratings of a familiar target individual; this can be ascertained by evaluating inter-rater reliability (Gosling, 2001). In addition, due to the definition of personality as "temporally stable patterns of affect, cognition, and behaviour" (Gosling, 2008), repeated testing of the same adult individual by the same rater using a personality assessment tool should yield consistent scores (Dingemanse and Wright, 2020). This is termed test-retest reliability. Therefore, a personality assessment tool can be considered to produce valid and reliable personality data if its internal consistency, concurrent validity, inter-rater reliability and test-retest reliability have been established.

Due to the non-verbal nature of animals, self-report using psychometric questionnaires cannot be used to gather animal personality data. The two main methods used to explore personality in animal models rely on either behavioural coding or trait ratings by familiar human raters (Gosling, 2001). Behavioural coding is generally undertaken while exposing the target animals to standardised experimental tests linked to the trait(s) of interest (Gosling, 2001). Some tests may be used across a wide range of species; for instance, exposure to a novel stimulus is widely used to measure boldness or exploration (Réale et al., 2007; Gosling, 2001). Others may take into account species-specific behaviour, such as the home-cage intruder test (Henry and Stephens, 1977) and tonic immobility back test (Hessing et al., 1994) used to assess coping style in rodents and piglets, respectively. Behavioural coding is considered advantageous by some authors as it results in the collection of objective data, although this view is disputed (Gosling, 2001). Indeed, standardised tests may be used to record purely objective parameters, such as latency to respond to the test stimulus or distance travelled (Réale et al., 2007). However, the frequency or duration of behaviour may also be recorded instead (Vazire et al., 2007). In this case, the precision of the definition of the traits or behaviour coded highly impacts on the objectivity of the data, as most definitions may require a degree of interpretation by the observer that may introduce a level of inter-rater variability (Gosling, 2001). In addition, this method is also limited as the standardised tests constitute only a point in time measure. This is problematic, as behaviour is influenced by a number of parameters such as circadian or circannual rhythms, or environmental disruptions, which may confound the response to a point-in-time tests (Vazire et al., 2007). Therefore, in order to truly represent personality, testing must be repeated over time and across situations (Gosling, 2001). Furthermore, the number of traits that can be evaluated using a single test is limited, meaning that full personality profiles can only be obtained by using a battery of tests (e.g. Lansade et al., 2016). Finally, in order to be accurate, behaviour coding must be conducted by trained research professionals (Vazire et al., 2007). Behavioural coding is therefore limited by logistical considerations and access to an expert workforce, and may not capture the expression of the target animal’s full behavioural repertoire.

In contrast to behaviour rating, subjective trait rating relies on gathering the impressions of a human familiar with the animal, usually a caregiver, using a rating system (Gosling, 2001). Unlike the behaviour coding method, this allows access to an aggregate view of the animal's behavioural tendencies over time and across situations (Gosling, 2001), with the observer providing a rating on the basis of the sum of their interactions with a familiar animal (e.g. Stevenson-Hinde, Stillwell-Barnes and Zunz, 1980). In addition, this method is not constrained by the same workforce limitations as behaviour codings, as non-trained raters can be used as long as they are familiar with the target individual and a variety of other individuals of the same species (Vazire et al., 2007). Finally, the subjective element of this method does not detract from its reliability (Vazire et al., 2007). Indeed, Gosling's review identified that trait-rating studies showed equal, if not better reliability as behavioural coding, as well as concurrent validity (Gosling, 2001). Therefore, subjective trait rating appears to be most appropriate way of obtaining personality data from a large sample of animals, where logistical constraints such as time and availability of research personnel apply. In order to produce meaningful measures of personality, the questionnaire instruments used to record subjective ratings should fulfil the same criteria of validity and reliability as human personality questionnaires (Gosling and Vazire, 2002; Taylor and Mills, 2006). Therefore, animal personality questionnaires should be constructed using the same psychometric approach as outlined above for human personality questionnaires (Simms and Watson, 2007), rather than relying on an ad hoc selection of traits. In addition, prior to its use for research or applied purposes, the internal consistency, concurrent validity, inter-rater reliability and test-retest reliability of an animal personality questionnaire should be established (Gosling and Vazire, 2002).

1.3 NEUROPHYSIOLOGICAL CORRELATES OF PERSONALITY

1.3.1 NEUROPHYSIOLOGICAL CORRELATES OF THE FIVE FACTOR MODEL

The main focus of early personality research was to establish a structure of human personality, rather than explore its driving mechanisms (Réale et al., 2007). Nevertheless, early works formulated theoretical frameworks of the neurobiology underlying personality factors of particular interest, such as Neuroticism and Extraversion (Eysenck, 1983; Depue and Collins, 1999). These theoretical models were used as a basis to explore links between personality factors and putative drivers of personality, traditionally through the use of physiological (reviewed in e.g. Ormel et al., 2013) or genetic parameters (reviewed in e.g. Munafò, 2009). While some of the systems targeted by those studies have since been identified as correlates, rather than drivers, of personality (Ormel et al., 2013; de Boer, Buwalda and Koolhaas, 2017), these studies nevertheless constitute a valuable information base to describe the physiological profile of personality types. In more recent years, the development of brain imaging techniques has enabled researchers to directly explore links between personality factors and brain anatomy and functioning (DeYoung and Gray, 2009). This new discipline, termed personality neuroscience, is leading to the development of a clearer model of the neural basis of personality factors (DeYoung et al., 2010).

Early theoretical frameworks of the biological basis of Neuroticism linked this factor with the biological systems associated with sensitivity to threats and punishment (Eysenck, 1983) and lower thresholds of activation in the sympathetic nervous system and limbic system (reviewed in e.g. Ormel et al., 2013). In particular, Neuroticism appears mediated by an increased sensitivity of the amygdala to negative stimuli (Ormel et al., 2013; DeYoung et al., 2010). Due to the regulatory effect of the amygdala on the autonomic nervous system and hypothalamic–pituitary–adrenal (HPA) axis, this increased sensitivity of the amygdala was hypothesised to translate to increased autonomic and HPA responses to negative stimuli (Ormel et al., 2013). Neural imagery studies appear to confirm the link between Neuroticism and increased activity in the amygdala (Ormel et al., 2013). However, empirical studies exploring links between Neuroticism and peripheral measures of autonomic and HPA axis reactivity only partially reflect the subsequent links hypothesised with these systems (Ormel et al., 2013).

A meta-analysis of studies carried out over 30 years investigating links between Neuroticism and responses to laboratory induced stress in healthy populations supports links between Neuroticism and cardiac parameters (Chida and Hamer, 2008). In contradiction with theoretical models (Eysenck, 1983), high Neuroticism is significantly associated with a decrease in heart rate and blood pressure reactivity to stressors (Chida and Hamer, 2008).Neuroticism is also associated with poorer cardiovascular recovery after exposure to the stressors had ceased (Chida and Hamer, 2008). However, Neuroticism was not significantly associated with sympathetic nervous system reactivity, and only tended to be associated with decreased parasympathetic nervous system reactivity (Chida and Hamer, 2008). Similarly, although theoretical frameworks suggest a positive link between Neuroticism and HPA axis reactivity, this is not supported by the majority of empirical studies available. Some studies link Neuroticism with decreased HPA axis reactivity (e.g. Phillips et al., 2005; Oswald et al., 2006); this surprising finding may be explained as a potential downregulation of the HPA axis in Neuroticism to protect against the negative impact of repeated activation (Ormel et al., 2013). However, meta-analyses and synthetic reviews show that most studies do not find a significant relationship between Neuroticism and HPA axis activity (Chida and Steptoe, 2009) or reactivity (Chida and Hamer, 2008; Ormel et al., 2013). Nonetheless, this finding is still the subject of investigation. Indeed, it has been suggested that null findings in multiple studies may be due to methodological issues, in particular around sampling times and failure to take into account potential confounding factors, rather than a true independence of Neuroticism and HPA axis reactivity (Ormel et al., 2013). Therefore, although the link between Neuroticism and amygdala reactivity has been confirmed (Ormel et al., 2013), the subsequent links between Neuroticism and the reactivity of systems involved in the physiological stress response are only partially supported in humans (Chida and Hamer, 2008). The main correlate of Neuroticism at the peripheral level appears to be decreased cardiovascular reactivity to stressors, as measured by heart rate and blood pressure (Chida and Hamer, 2008). There is also partial support for links with decreased reactivity of the parasympathetic nervous system, as measured by heart rate variability, and decreased HPA axis reactivity, as measured through cortisol levels (Chida and Hamer, 2008).

Extraversion is associated with positive affect (Depue and Collins, 1999). Early theoretical frameworks have therefore linked this personality factor with the biological systems underlying reward sensitivity and positive incentive motivation (Depue and Collins, 1999). Given the centrality of dopaminergic networks in incentive motivation, dopaminergic networks were also hypothesised as underlying drivers of Extraversion, with a particular emphasis on the ventral tegmental area and its dopaminergic projections onto the nucleus accumbens, the amygdala and the orbitofrontal cortex (Depue and Collins, 1999). Although a firm conclusion has not yet been reached regarding the exact neural basis of Extraversion, a number of strands of empirical studies confirm the link between Extraversion and dopaminergic systems. For instance, associations have been found between Extraversion and genes involved in the dopaminergic system (e.g. reviewed in Munafò, 2009). In addition, functional brain imaging has revealed positive associations between Extraversion and activity in the brain circuitry linked with reward processing in response to positive stimuli (DeYoung and Gray, 2009). The regions highlighted include the medial orbito-frontal cortex, nucleus accumbens, amygdala and striatum (DeYoung and Gray, 2009), all structures that form part of the mesocorticolimbic dopaminergic system and receive projection of dopaminergic neurons from the ventral tegmental area (Cabib, 2006). Therefore, empirical evidence appears to support a positive link between Extraversion and dopaminergic activity in the brain circuits linked with reward processing. However, it should be noted that the dopaminergic model may not drive all facets of Extraversion (Depue and Morrone-Strupinsky, 2005). Indeed, it has been argued that while dopamine may underlie facets related to motivational drive, such as Assertiveness and Activity (Depue and Collins, 1999), facets related to affiliation such as sociability may instead be driven by the endogenous opioid system, with oxytocin and vasopressin as the main mediators (Depue and Morrone-Strupinsky, 2005).

Agreeableness, Openness and Conscientiousness, the remaining three dimensions in the Five Factor Model, cannot be mapped as clearly onto biological systems; therefore, to date, theoretical frameworks of their biological bases are lacking (DeYoung et al., 2010). Nonetheless, neurotransmitters (DeYoung and Gray, 2009) and brain regions (DeYoung et al., 2010) of interest have been identified. Deyoung (2006) presents evidence that the Five Factor model factors can be grouped onto two meta-traits, referred to as “Stability” and “Plasticity”. In a factorial analysis, Emotional Stability (the opposite pole of Neuroticism), Agreeableness and Conscientiousness load onto the meta-trait Stability , while Extraversion and Openness load onto the meta-trait Plasticity (Deyoung, 2006). These two traits are uncorrelated and hypothesised to be driven by separate biological substrates (Deyoung, 2006): Stability is related to differences in serotonergic functioning (DeYoung and Gray, 2009), while Plasticity is dopamine driven (DeYoung and Gray, 2009; DeYoung, 2013). Therefore, it has been hypothesised that differences in Agreeableness and Conscientiousness are linked with differences in serotonin network functioning, while Openness is linked to differences in dopamine function (DeYoung and Gray, 2009). In addition, brain imaging studies suggest that Agreeableness may be linked with anatomical differences in brain regions associated with empathy and the interpretation of other people's actions and beliefs (DeYoung et al., 2010). By contrast, emerging evidence suggests Conscientiousness may be linked with brain regions associated with self-regulation (DeYoung et al., 2010).

In summary, while the exact biological basis of personality remains to be established, theoretical frameworks of the underlying drivers of personality have now received ample support from empirical studies. The neurophysiological basis of the Five Factor Model traits is becoming better established; in particular, clear positive links have been confirmed between Extraversion and dopaminergic activity. In addition, physiological correlates of personality have also been identified at the peripheral level. In particular, Neuroticism has been linked with decreased cardiovascular and parasympathetic reactivity to stressors.

1.3.2 NEUROPHYSIOLOGICAL CORRELATES OF PERSONALITY IN ANIMAL MODELS

In contrast to the relatively well-established model of neurophysiological bases and correlates of personality in humans reviewed in Section 1.3.1, animal models of personality tend to focus on behavioural expressions of personality (Gosling, 2001) and often do not comprise detailed information on physiological correlates (Rankins and Wickens, 2020). An exception to this is the coping style framework, in which physiological correlates of coping style have been extensively described, along with potential neurophysiological drivers of differences in behaviour (reviewed in e.g. Koolhaas et al., 1999, 2010; de Boer, Buwalda and Koolhaas, 2017).

At the peripheral level, the two coping styles are associated with differential patterns of baseline function and reactivity of the autonomic nervous system, as well as neuroendocrine systems such as the hypothalamic-pituitary-adrenocortical (HPA) axis (reviewed in e.g. Koolhaas et al., 2010; de Boer, Buwalda and Koolhaas, 2017). Proactive and reactive individuals show comparable baseline levels of plasma catecholamine (adrenaline, noradrenaline), suggesting comparable baseline activity of the sympathetic nervous system (Koolhaas et al., 2010). By contrast, studies using high and low aggressive rodent strains in laboratory settings have demonstrated an increased adrenal medulla and sympathetic reactivity to stressors in proactive individuals (de Boer, Buwalda and Koolhaas, 2017). This is evidenced by an increased plasma catecholamine response (Fokkema, Koolhaas and van der Gugten, 1995; Carnevali et al., 2013), increased incidence of tachyarrhythmia (fast heart rate) (Carnevali et al., 2013), increased blood pressure reactivity (Fokkema, Koolhaas and van der Gugten, 1995), and increased respiratory rate (Carnevali, Nalivaiko and Sgoifo, 2014) during exposure to a stressor. By contrast, reactive individuals show increased parasympathetic reactivity to stressors, marked by a higher heart rate variability response to stressors (Sgoifo, Carnevali and Grippo, 2014). This finding has been replicated in a number of species, such as pigs (Hessing et al., 1994) and laying hens (Korte et al., 1997; Korte, Ruesink and Blokhuis, 1998). Therefore, proactive individuals show increased sympathetic reactivity to stressors, while reactive individuals show increased parasympathetic reactivity.

The two coping styles are also associated with differential patterns of baseline function and reactivity of neuroendocrine systems such as the hypothalamic-pituitary-adrenocortical (HPA) axis in some species (reviewed in e.g. Koolhaas et al., 2010; de Boer, Buwalda and Koolhaas, 2017). Results in rodents suggest that reactive individuals may show both higher baseline activity and higher reactivity of the HPA axis to stressors (Korte et al., 1992; Veenema et al., 2003; Koolhaas et al., 2010). Similar findings have been documented in a range of mammalian and avian species such as pigs (Hessing et al., 1994), eastern chipmunks (Martin and Réale, 2008), laying hens (Korte et al., 1997) and great tits (Carere et al., 2003). However, there appears to be some variations between species in the strength of the association between activity of the HPA axis and coping style (briefly reviewed in Koolhaas et al., 2010). Therefore, reactive individuals present higher HPA axis reactivity to stressors, although this finding may not be generalisable across species.

Koolhaas and collaborators (Koolhaas et al., 2010; de Boer, Buwalda and Koolhaas, 2017) argue that that these differences in nervous and neuroendocrine reactivity might be correlates, rather than causes, of the individual differences in behaviour patterns observed in coping style. Behavioural and physiological manifestations of coping styles are thought to share common underlying neurophysiological drivers (de Boer, Buwalda and Koolhaas, 2017). A causal model of coping styles has been proposed, identifying neurophysiological drivers of coping style at the centre level (de Boer, Buwalda and Koolhaas, 2017). This model is based on a subsequent refinement of the coping style framework using three axes to explain differences in coping style: reward sensitivity, executive control, and emotional arousal, which represents physiological stress sensitivity (de Boer, Buwalda and Koolhaas, 2017). These three axes are proposed to map onto separate brain networks, with neurochemical differences in those networks explaining differences in behaviour and peripheral physiology (de Boer, Buwalda and Koolhaas, 2017).

Similar to Extraversion in the Five Factor Model, the reward sensitivity axis is thought to map onto the dopamine pathway linking the ventral tegmental area to the nucleus accumbens (de Boer, Buwalda and Koolhaas, 2017), with proactive individuals showing higher striatal dopamine levels (Benus et al., 1991). In addition, in line with the hypothesis that the human meta-trait “Stability” may be driven by serotoninergic systems (DeYoung and Gray, 2009), the executive control axis of the coping style model is thought to be driven by differences in serotonin levels in the prefrontal cortex (de Boer, Buwalda and Koolhaas, 2017). In line with their high levels of aggression and reduced behavioural flexibility, proactive individuals show reduced serotonin levels in the prefrontal cortex compared to reactive individuals (Korte et al., 1996; De Boer and Koolhaas, 2005; Caramaschi, de Boer and Koolhaas, 2007). Finally, emotional arousal has been mapped onto the amygdala-hypothalamus-periaquaductal pathway (de Boer, Buwalda and Koolhaas, 2017). Precise causal neurochemical mechanisms have not yet been identified to explain individual differences in emotional arousal. However, differences in vasopressin/oxytocin balance within this pathway, and specifically at the level of the amygdala, are hypothesised as an underlying driver of differences in emotional arousal (Everts and Koolhaas, 1999; Calcagnoli et al., 2014).

In summary, a causal model is available for the coping style framework based on brain networks and neurotransmitters common to those identified as putative biological drivers of the human Five Factors (de Boer, Buwalda and Koolhaas, 2017). This is consistent with findings in comparative personality psychology suggesting that basic personality dimensions are highly conserved across species (Gosling, 2001; Gosling and John, 1999). In addition, physiological correlates of coping style have also been identified (Koolhaas et al., 2010). These physiological correlates are largely in line with those proposed as part of the theoretical framework of Neuroticism (Ormel et al., 2013). However, relationships between physiology and coping styles seem to emerge more clearly from animal studies, perhaps due to the use of artificially selected strains rather than a continuous population (Réale et al., 2007). These findings appear largely conserved across species and may therefore be used as a reasonable base to propose hypotheses when exploring neurophysiological correlates of personality in a new species.

1.4 EQUINE PERSONALITY

1.4.1 APPLIED RELEVANCE OF EQUINE PERSONALITY MEASUREMENT

Equine personality is assessed in the context of research (e.g. Morris, Gale and Howe, 2002; Roberts et al., 2016), but is also considered a highly relevant characteristic of a horse at all levels of the industry (e.g. Graf, König von Borstel and Gauly, 2013; König von Borstel et al., 2013). Personality is the most universally valued attribute of a horse for relatively novice horse owners with a Pony Club membership, and is the most important purchase criterion ahead of health characteristics such as soundness (Buckley, Dunn and More, 2004). The two personality traits "character" and "temperament" are identified as the most important traits to select on in a list of breeding goals, ahead of rideability and sporting attributes such as the quality of gaits, by a large scale international sample of professional, competition and leisure riders and breeders (Graf, König von Borstel and Gauly, 2013). In addition, 100% of German breed judges and professional test riders who answered a survey on personality testing in the horse deemed personality an important or very important factor in rideability, quality of horse-human relationship and performance (König von Borstel et al., 2013). Finally, an economic evaluation of objective personality assessment has revealed equestrians would be willing to pay for an objective evaluation of equine personality, in particular for breeding stock, and that favourably assessed horses would increase in value (Graf et al., 2013). These studies therefore highlight the importance of equine personality to a wide range of equestrians, ranging from professionals to amateurs.

In line with the perceived importance of personality, equestrians express clear preferences for some equine personality types. A quiet, calm and tolerant personality is considered desirable in horses used to teach novice riders such as riding school horses (Odberg and Bouissou, 1999), horses used in the Pony Club (Buckley, Dunn and More, 2004) or therapy horses (Anderson et al., 1999; Grandin, Fine and Bowers, 2010). In this context, compliance is seen as an indicator of good performance (Buckley, Dunn and More, 2004). In addition, although it has been hypothesised that increased emotional reactivity may be beneficial in some disciplines (e.g. racing: McBride and Mills, 2012),this preference appears to extend to competitive riders. Most respondents in a large-scale survey, including professional and competition riders, place large amounts of importance on traits such as even-temperedness, attention sensitivity to riders' aids, sociability towards humans, ease of habitutation to new surroundings, and behaviour while grooming and tacking up (Graf, König von Borstel and Gauly, 2013). Similarly, studies explicitly exploring desired traits in both leisure and sport horses consistently highlight the importance of compliance, ease of training, predictability, sociability towards humans and horses and low fearfulness, irrespective of the background and equitation skill level of respondents (Górecka-Bruzda et al., 2011; Suwała et al., 2016). The majority of equestrians, including competitive riders, disagree that sports horses need a difficult personality (Graf, König von Borstel and Gauly, 2013) and do not desire a challenging horse (Górecka-Bruzda et al., 2011). In accordance with this, eventers at the highest levels of the sport highlight traits such as emotional stability, boldness and compliance with rider demands as key traits of a successful eventing horse (Wipper, 2000). Therefore, equestrians at all levels of the industry appear to place value on horses who show predictable, compliant responses and limited behavioural reactivity.

Informal personality assessment is used widely in the industry to select horses for roles, according to their perceived compatibility with the personality profile thought to be required for the role. For instance, the traditional recruitment process for police horses includes a subjective assessment by experienced police riders or trainers to assess their suitability (e.g. Munsters et al., 2013; Pierard, McGreevy and Geers, 2017; Norton et al., 2018). Selection on the basis of informal personality assessments is also documented or recommended in the literature for therapy horses (Grandin, Fine and Bowers, 2010; Anderson et al., 1999), riding school (König von Borstel, 2013; Odberg and Bouissou, 1999) and Pony Club horses (Buckley, Dunn and More, 2004), vaulting horses (Hausberger, Muller and Lunel, 2011), and competition horses such as eventers (Wipper, 2000). However, these judgements are usually not evidenced-based and may lack objectivity (König von Borstel et al., 2013). This may result in behavioural problems due to a mismatch between the horse and its assigned rider or role, leading to wastage (Wolframm, Gerardus and Meulenbroek, 2012; Munsters et al., 2013b).

Although personality is recognised as a clear driver of performance (McBride and Mills, 2012; Roberts et al., 2016) on par with physical abilities and conformation (Visser et al., 2003a;

Suwała et al., 2016), relatively few studies have been conducted to support this view to date. Nonetheless, where associations are investigated, personality traits are often significantly linked with aspects of performance. For instance, emotionality traits can be used to predict show-jumping performance (Visser et al., 2003a; Lansade et al., 2016). In addition, equine emotionality may also negatively influence the quality of the horse-human relationship (Visser et al., 2008), which is considered a key component of optimal performance (Wipper, 2000). Finally, an impact of personality on cognitive style has also tentatively been documented, with a particular impact of the trait "fearfulness". More "fearful" horses may be more prone to habit formation, as they have been found to be less sensitive to a contingency degradation protocol when learning an instrumental task (Lansade et al., 2017), and to be more resistant to extinction in an extinction paradigm (Valenchon et al., 2013). Therefore, while research remains ongoing, current evidence suggests that personality has an impact on a number of components of equine performance, and that selection on the basis of personality may be reflected by an impact on these correlates.

Emerging findings on the personality characteristics of successful performance horses highlight the importance of systematically investigating popular assumptions of the characteristics required for a particular role. For instance, in line with the popular idea that a horse with a more active flight response might possess more speed, it has been hypothesised that increased "flightiness", or emotional reactivity, might be desirable in a racehorse (McBride and Mills, 2012). This is supported to a degree by the finding that Thoroughbreds as a breed score higher on personality components "Anxiousness" and "Excitability" than cold-blooded horse and pony breeds such as Irish draughts or Highland and Shetland ponies (Lloyd et al., 2008). However, when flightiness is investigated within the breed standard of Thoroughbreds, empirical data suggests that higher extremes of flightiness might be a hindrance, rather than a competitive advantage. Indeed, highly reactive behaviour ahead of the race has been significantly linked with poor performance (Hutson and Haskell, 1997). In addition, Thoroughbreds rated as “temperamental” rather than “calm”, using a scale combining the “anxiousness” and “excitability” factors from (Lloyd et al., 2007), show no increase in speed but have significantly higher physiological stress responses to training and take longer to return to baseline physiological state, suggesting they may be more fatigued and need longer to recover from an exercise bout (Bohák et al., 2017). Therefore, there appears to be an optimum level of flightiness required for use as a racehorse, under or above which performance is hindered. This example illustrates the importance of considering not only behavioural, but also cognitive and physiological correlates of personality when drawing an evidence-based profile of desirable personality traits for a role.

1.4.2 MEASUREMENT TOOLS

As in other animals, personality has been assessed in the horse using either objective behavioural coding during standardised tests, or subjective trait-rating by familiar observers (recently reviewed in Rankins and Wickens, 2020). Standardised tests used for behavioural coding echo those used in other species (Rankins and Wickens, 2020), although some are adapted to account for the domestic horses’ specific behavioural repertoire and management (e.g. Wolff, Hausberger and Le Scolan, 1997).

The majority of tests measure fear or reactivity responses in a variety of contexts (Rankins and Wickens, 2020). Novelty tests aim to document trait-neophobia and exploration by presenting the horse with a previously unknown stimulus (generally an object), with free choice to explore (e.g. Lansade, Bouissou and Erhard, 2008; Ijichi et al., 2013). By contrast, bridge tests incorporate a handling dimension and require the horse to be led over a novel surface (e.g. Wolff, Hausberger and Le Scolan, 1997; Ijichi et al., 2013). Startle or reactivity tests measure reactivity by exposing the horse to a sudden, usually visual, stimulus (e.g. Lansade, Bouissou and Erhard, 2008; Ijichi et al., 2013).

Additionally, other tests may be used to measure other dimensions of personality. For instance, sensitivity to touch may be investigated using von Frey filaments (Lansade, Pichard and Leconte, 2008), reactivity to humans may be measured using passive or active human tests (Lansade and Bouissou, 2008), and sensitivity to social isolation may be measured using an open arena test in a familiar arena (Wolff, Hausberger and Le Scolan, 1997). Due to its perceived objectivity, behavioural coding during standardised tests has so far been the preferred method used to implement formal personality assessment in the field, and in particular in breed shows or breeding selections (Lansade et al., 2016; Graf, König von Borstel and Gauly, 2014). Batteries of tests investigating different dimensions of personality have been developed for use in the field (Graf, König von Borstel and Gauly, 2014; Lansade et al., 2016). In both cases, the main dimensions assessed are related to behavioural reactivity to frightening stimulus and tactile sensitivity. The use of behavioural tests is preferred in an applied context such as breeding selection as they are perceived to be more objective and are not as easily manipulated for commercial gain as subjective trait-ratings (Graf, König von Borstel and Gauly, 2014). However, these tests do not typically result in an exhaustive overview of equine personality and do not clearly give rise to a model of personality compatible with comparative findings in other species.

A number of subjective trait-based assessment tools have also been used to investigate equine personality in a research context (reviewed in Rankins and Wickens, 2020). While some studies simply assessed personality at the behaviour or trait level and did not investigate factor structure (e.g. Mills, 1998; Anderson et al., 1999; Seaman, Davidson and Waran, 2002), most used factorial analysis in order to propose a model of equine personality. Early questionnaires appear to have been developed using an ad hoc approach to item selection, with little justification offered behind the selection of traits used (Anderson et al., 1999; Seaman, Davidson and Waran, 2002; Visser et al., 2003b; Momozawa et al., 2005a). Perhaps as a result of these limitations in development, most early questionnaires were shown to have limited concurrent validity (Anderson et al., 1999; Seaman, Davidson and Waran, 2002; Visser et al., 2003b) and inter-rater reliability (Anderson et al., 1999; Visser et al., 2003b). However, subsequent questionnaires (McGrogan, Hutchison and King, 2008; Ijichi et al., 2013; Lloyd et al., 2007) were developed using a psychometric approach in line with recommendations in the psychology literature (Simms and Watson, 2007; Gosling and Vazire, 2002). For these questionnaires, a pool of prospective items of interest thought to describe equine personality was selected, either on the basis of existing lists of traits used in other species (e.g. Stevenson-Hinde and Zunz, 1978; used as a basis for Lloyd et al., 2007), or in consultation with experts recruited from the equine industry (McGrogan, Hutchison and King, 2008; Creighton, personal communication), with the intention to produce an exhaustive list of personality descriptors for the species. The reliability of items was then evaluated on the basis of agreement between multiple raters for each target horse, and only reliable items were included in a factorial analysis (Lloyd et al., 2007; McGrogan, Hutchison and King, 2008; Creighton, personal communication).

Psychometrically developed equine personality questionnaires generally perform well when tested against the criteria for a valid and reliable personality assessment tool laid out by Gosling and Vazire (2002) and Dingemanse and Wright (2020). For instance, McGrogan, Hutchison and King (2008) report high values of internal consistency and inter-rater reliability, on par with benchmarks set by human personality inventories (e.g. Mccrae and Costa, 1987), for the three factors extracted from their questionnaire. In addition, both Lloyd et al. (2007) and Ijichi et al. (2013) have evaluated the concurrent validity of their questionnaire and found correlations between personality scores and either naturally occurring behaviour (Lloyd et al., 2007) or responses to behavioural tests designed to measure similar traits (Ijichi et al., 2013). Further, predictive validity was also demonstrated for the Equine Personality Test (Ijichi et al., 2013), as it was shown to predict pain expression in the horse (Ijichi, Collins and Elwood, 2014). The use of a psychometric approach in developing equine personality questionnaires therefore appears to give rise to more robust assessment tools. However, it should be noted that to date, none of these three questionnaires have been validated against all four criteria set out by Gosling and Vazire (2002) and Dingemanse and Wright (2020). In particular, no equine personality questionnaire has yet been assessed for test-retest reliability.

1.4.3 EMERGING MODEL OF EQUINE PERSONALITY

Several factor structures of equine personality have been proposed, comprising of between two (Visser et al., 2003) and nine (Roberts et al., 2016) personality dimensions. A four to six factor solution appears most common (Ijichi et al., 2013; McGrogan, Hutchison and King, 2008; Lloyd et al., 2007; Momozawa et al., 2005a; Morris, Gale and Howe, 2002). However, there is little consensus to date on a model of equine personality, with varied dimensions emerging from the structure analyses in the studies above. This is most likely due in part to differences in methodologies used to develop the questionnaires, as well as a lack of continuity in language between studies when naming factors emerging from the factorial analysis. In addition, the factor structure emerging from any trait-based personality assessment tool is in part dependent on the range of traits included in the questionnaire: dimensions that contain traits not included in the questionnaire cannot emerge from factor analysis even if they are present in the species (Gosling and John, 1999). Nonetheless, when factors are examined at the trait level, some elements of common structure compatible with findings in other species emerge.

To the best of the author’s knowledge, all equine personality questionnaires developed to date yield a factor consistent with sensitivity or reactivity to stressors, comparable to FFM Neuroticism. These factors have been termed Response to the environment (Visser et al., 2003), Anxiety (Momozawa et al., 2005a; Roberts et al., 2016), Anxiousness (Lloyd et al., 2007), or Neuroticism (Morris, Gale and Howe, 2002; McGrogan, Hutchison and King, 2008; Ijichi et al., 2013; Olsen and Klemetsdal, 2017). This factor comprises traits such as “spooky”, “nervous”, “fearful”, or “tense” (Visser et al., 2003b; Momozawa et al., 2005a; Lloyd et al., 2007; Ijichi et al., 2013) and is predictive of the intensity of the startle response to a sudden stimulus for the only questionnaire validated against behavioural tests (Ijichi et al., 2013). Perhaps because Neuroticism predicts responses that may be dangerous to human handlers (Ijichi et al., 2013), this factor is extracted consistently by personality questionnaires (Momozawa et al., 2005a), comprises a large number of traits, and is generally rated reliably by observers (Morris, Gale and Howe, 2002).

In addition to Neuroticism-like factors, a majority of questionnaires extract a factor comparable to FFM Agreeableness. This factor relates to traits such as compliance, cooperation, friendliness, and non-aggressiveness, and is labelled affability (Momozawa et al., 2005a) or Agreeableness (Morris, Gale and Howe, 2002; Ijichi et al., 2013; McGrogan, Hutchison and King, 2008; Olsen and Klemetsdal, 2017). In accordance with the high relative importance given to a compliant personality by equestrians, this factor generally comprises a large number of items in questionnaires based on traits suggested by industry experts (Ijichi et al., 2013; McGrogan, Hutchison and King, 2008). Interestingly, Lloyd et al. (2007), whose Horse Personality Questionnaire is based in majority on an adjective list designed for use in primates (Stevenson-Hinde and Zunz, 1978; Stevenson-Hinde, Stillwell-Barnes and Zunz, 1980), did not extract a specific Agreeableness-like component. However, in a subsequent revision of the Horse Personality Questionnaire including an additional 13 traits suggested by a focus group of equine professionals, a separate Agreeableness factor emerged (Olsen and Klemetsdal, 2017). Two questionnaires yield a factor likened to the Five Factor Model Conscientiousness (Morris, Gale and Howe, 2002; Olsen and Klemetsdal, 2017). This is surprising as Conscientiousness is rarely identified in non-human species (Gosling and John, 1999), when understood as a measure of impulse control (Mccrae and Costa, 1987). However, closer inspection at the trait level suggests that these two factors reflect willingness to work and compliance rather than impulse control. These factors may therefore arguably represent facets of Agreeableness instead. Indeed, several of the traits included in the Conscientiousness factor suggested by Olsen and Klemetsdal (2017) load on Agreeableness factors in other models: cooperative (McGrogan, Hutchison and King, 2008), (non-)stubborn (Ijichi et al., 2013; McGrogan, Hutchison and King, 2008), and willing (Ijichi et al., 2013).

In addition to Neuroticism and Agreeableness, equine personality models often yield a factor consistent with FFM Extraversion, or several factors reflecting separate facets of Extraversion, such as Assertiveness, Activity or Sociability. Two questionnaires use the term Extraversion directly (Ijichi et al., 2013; McGrogan, Hutchison and King, 2008). However, one factor appears to reflect in majority Assertiveness, with traits such as “bold”, “competitive” and “not timid” (McGrogan, Hutchison and King, 2008), while the other seems to reflect in majority Activity, with traits such as “excitable”, “spirited” and “active” (Ijichi et al., 2013). In addition, three studies describe an “Excitability” factor that appears consistent with the Activity facet of Extraversion (Lloyd et al., 2007; Olsen and Klemetsdal, 2017; Roberts et al., 2016), including traits such as “active”, “excitable” and “impulsive” (Olsen and Klemetsdal, 2017). Interestingly, most studies that include items relating specifically to relationships with conspecifics yield separate factors related to sociability or gregariousness (Ijichi et al., 2013; Roberts et al., 2016; Lloyd et al., 2007). Sociability is generally considered a key feature of human Extraversion (Mccrae and Costa, 1987). However, some authors have suggested that sociability and surgency (i.e. assertiveness and activity) may constitute two separate factors, rather than two factors of Extraversion (Mccrae and Costa, 1987). In line with this, the definition of the conserved Extraversion factor evidenced across multiple species does not include a social dimension, and is closer to the definition of boldness (Gosling and John, 1999). Accordingly, in horses, sociability or gregariousness appears to emerge as a separate factor from Extraversion (Ijichi et al., 2013; Roberts et al., 2016; Lloyd et al., 2007).

Finally, other factors emerge more occasionally. These include Inquisitiveness (Lloyd et al., 2007; Roberts et al., 2016), a factor that includes traits such as “curious”, “opportunistic” and “playful”. This factor may reflect the conserved Openness factor described by Gosling and John (1999), which is centred around curiosity towards novelty and playfulness. As argued by Gosling and John (1999), the inconsistent evidence for this factor may be due to the fact that some questionnaires do not include items relating to this factor (Momozawa et al., 2005a). Indeed, these traits may be of less importance to equestrians than those relating to Neuroticism and Agreeableness and may therefore not be suggested by panels of experts. However, it should be noted that when present, items that appear conceptually close to Openness do not always form a separate factor. For instance, “Adventurous/habitual” loads onto Extraversion in the Equine Personality Test (Ijichi et al., 2013). Dominance, another factor identified across species by Gosling and John (1999), also emerges in the two studies based on Stevenson-Hinde and Zunz's (1978) adjective list (Lloyd et al., 2007; Olsen and Klemetsdal, 2017).

When considering the available equine personality questionnaires, it appears that the conclusions drawn by Gosling and John (1999) also apply to horses, with empirical evidence supporting the existence of three main personality factors consistent with FFM Neuroticism, FFM Agreeableness, and the facets of FFM Extraversion linked with activity. In addition, gregariousness or sociability, as well as dominance, frequently emerge as separate factors. By contrast, there is inconsistent support for an additional Openness factor. Given the calls for a unified language in animal personality studies (Gosling, 2001; Réale et al., 2007), it appears advantageous to base future research on equine personality questionnaires that explicitly align with this conserved structure of animal personality (Gosling and John, 1999), such as the Equine Personality Test (EPT) developed by Creighton (Creighton, personal communication; full text of the questionnaire published in Ijichi et al., 2013). From an applied point of view, the desired personality profile of an emotionally stable, compliant horse (Graf, König von Borstel and Gauly, 2013; Suwała et al., 2016; Górecka-Bruzda et al., 2011) has yet to be formally mapped onto this emerging model. However, trait level analysis suggests that the desired characteristics would translate as low Neuroticism and high Agreeableness. On those questionnaires measuring primarily the Assertiveness facet of Extraversion, it is likely that moderate Extraversion would also be desired. It therefore appears likely that, where selection on the basis of personality is applied, it occurs on the basis of those highly conserved traits.

Although a model of equine personality is now emerging from fundamental research into equine personality structure, relatively little remains known about non-behavioural correlates of personality in the horse (Rankins and Wickens, 2020). In particular, links between personality and physiological correlates have only been explored very sporadically (Rankins and Wickens, 2020), and the biological substrates of personality have not yet been determined in the horse.

1.4.4 LINKS BETWEEN PHYSIOLOGY AND PERSONALITY DOCUMENTED IN THE HORSE

To date, no empirical evidence is available to directly link specific brain networks to personality dimensions in the horse. While a model of brain structures and networks relevant to equine behaviour has been proposed (McBride et al., 2017), much of the evidence is generalised from rodent studies rather than specific to the horse. However, a small body of genetic studies supports links between differences in neurotransmitter metabolism and equine personality (Momozawa et al., 2005b; Ninomiya et al., 2013; Momozawa et al., 2006; XiuJuan et al., 2015; Hori et al., 2016). The dopamine receptor D4 gene has been linked with the temperament traits curiosity and vigilance (Momozawa et al., 2005b), as well as the tendency to express behavioural frustration ahead of meal times (Ninomiya et al., 2013). Those traits appear in line with the notion of positive incentive motivation and drive to obtain reward that are hypothesised as the biological basis of Extraversion in humans, and with the "Plasticity" meta-trait (Depue and Collins, 1999; DeYoung, 2013). Studies investigating links between serotonin and equine personality yield more inconsistent results. Different haplotypes of the Equine Serotonin Transporter gene are not associated with Anxiety (Momozawa et al., 2006). However, a mutation in the same gene is associated with the traits nervousness, panic and timidity (XiuJuan et al., 2015). Different haplotypes of the serotonin receptor 1A are also associated with different levels of "tractability", as assessed through a questionnaire (Hori et al., 2016). Although this questionnaire was not developed using a psychometric approach and lacks validation, it aims to assess traits relating to behavioural reactivity and compliance (Hori et al., 2016). Therefore, despite some null results, the available evidence suggests a link between serotonin and equine personality traits linked with Neuroticism and Agreeableness. Genetic studies using a fully validated, psychometrically designed questionnaire are currently lacking, which limits the reliability of results. Nonetheless, current evidence suggests that serotonin and dopamine may be two highly relevant mediators of personality in the horse, in line with models proposed in humans (DeYoung and Gray, 2009) and rodents (de Boer, Buwalda and Koolhaas, 2017).

At the peripheral level, preliminary evidence suggests associations between personality and autonomic reactivity to stressors. Autonomic reactivity has been positively linked with temperament dimensions assessing sensitivity to stressors derived from subjective questionnaires (Visser et al., 2003b; Momozawa et al., 2003). Indeed, heart rate reactivity to an experimental challenge (novel object test) was shown to be positively correlated with responsiveness to the environment as assessed under saddle by a rider, while it was negatively correlated with attentiveness to the rider Visser et al. (2003). In addition, it was also shown to be positively correlated with a temperament factor labelled "anxiety", reflecting traits such as “nervousness”, “stubbornness” and “excitability” (Momozawa et al., 2003). However, other studies suggest that behavioural reactivity to stressors as measured using standardised tests may be independent from autonomic reactivity (Lansade, Bouissou and Erhard, 2008; Squibb et al., 2018) Preliminary evidence also suggests differences in HPA axis reactivity may exist between more or less behaviourally reactive horses, although the direction of effect reported varies (Fazio et al., 2013; Bohák et al., 2017). Bohák et al. (2017) report lower HPA axis responses to a bout of exercise in calm than temperamental Thoroughbreds. By contrast, Fazio et al. (2013) report higher HPA axis responses to transport stress in calm than nervous stallions. These inconsistent results echo those documented in studies investigating a potential link between Neuroticism and physiolocal reactivity to stressors in humans (reviewed in Ormel et al., 2013), and do not yet paint a clear picture of the potential relationship between equine personality and physiological parameters. Therefore, it is unclear whether tests designed to select horses on the basis of behavioural reactivity to stressors (Lansade et al., 2016; Pierard, McGreevy and Geers, 2017) also reflect physiological reactivity to those same stressors.

1.5 RATIONALE, AIMS AND OBJECTIVES

As a result of the last decades of research effort, a tentative model of equine personality coherent with findings in other species (Gosling and John, 1999) emerges (reviewed in Section 1.4.3). In addition, batteries of objective tests (Graf, König von Borstel and Gauly, 2014; Lansade et al., 2016) as well as subjective questionnaires (e.g. Ijichi et al., 2013) are available to select horses on the basis of their behavioural responses to challenges. In most cases, these tests aim to select horses who display compliant behaviour and limited behavioural reactivity to challenges (e.g. Lansade et al., 2016; Pierard, McGreevy and Geers, 2017). From an applied point of view, this is advantageous to owners and handlers as this is expected to lead to increases in handler safety and reduction in wastage linked with behavioural issues incompatible with the horse's role (Ijichi et al., 2013; Munsters et al., 2013a). However, in contrast to models developed in humans (Depue and Collins, 1999; Ormel et al., 2013) and rodents (de Boer, Buwalda and Koolhaas, 2017), very little research has investigated neurophysiological correlates of personality in the horse (Rankins and Wickens, 2020).As a consequence, there is currently only a very limited understanding of what, if any, physiological characteristics are being selected for when selecting for a compliant and emotionally stable personality type.

Gaining further understanding of this is crucial for a number of reasons. Firstly, behavioural reactivity to challenges may only give limited insight into the horse's subjective experience of the stressors and their ability to cope from a physiological point of view (Munsters et al., 2013b; Yarnell, Hall and Billett, 2013; Squibb et al., 2018). Therefore, the objective tests developed to select horses for roles on the basis of their personality may not be sufficient to safeguard welfare. Thus, from an applied point of view, it is crucial to establish whether these tests predict physiological, as well as behavioural, reactivity to stressors. In addition, striatal dopamine levels, one of the main neurophysiological correlates of personality identified in other species, have been linked with cognitive style and susceptibility to stereotypies in the horse (McBride and Hemmings, 2005; Hemmings, McBride and Hale, 2007; Roberts et al., 2015). Therefore, clarifying the relationship between personality and dopamine in the horse may provide insight into potential impacts of the selection of the basis of personality onto these related characteristics.

The overarching aim of this thesis was therefore to explore physiological correlates of equine personality, in order to determine whether a physiological profile of the different personality types could be established in line with findings in other species. Potential neurophysiological correlates chosen for this preliminary investigation were those that: had been identified as correlates of personality in human and/or animal models (reviewed in 1.3); were of applied relevance to equine welfare or performance as discussed above; and could be studied using strictly non-invasive methods. Therefore, at the peripheral level, cardiac, parasympathetic and HPA axis reactivity to stressors were evaluated, as well as chronic HPA axis activity. At the central level, striatal dopamine was considered. Details and justifications of assessment methods are provided in each data chapter.

The Equine Personality Test (EPT: Ijichi et al., 2013) was chosen as the equine personality assessment tool to use throughout this thesis. This questionnaire was developed by Creighton (Creighton, personal communatication) and published in full by (Ijichi et al., 2013) This choice was motivated by the psychometric methods used in developing the questionnaire (Creighton, personal communication; Ijichi et al., 2013) and the pre-existing integration of the EPT in a comparative model of personality (reviewed in 1.4.3). Indeed, three of the five personality factors assessed by the EPT are aligned with human personality factors from the Five Factor Model (Mccrae and Costa, 1987), which have also been identified in a wide range of non-human animal species (Gosling and John, 1999). In addition, previous work has also drawn parallels between personality as measured by the EPT and the coping style framework as developed by Koolhaas et al. (1999), with Neuroticism and Extraversion thought to represent stress reactivity and coping style (Ijichi, 2014; Ijichi et al., 2013; Ijichi, Collins and Elwood, 2014). This was considered advantageous as it facilitated the generation of hypotheses based on results in other human and non-human animal species.

While the validity of the EPT had previously been established (Ijichi et al., 2013), its reliability had not yet been confirmed in a peer-reviewed publication at the time of starting this work. The first aim of this thesis, addressed in Chapter 3, was therefore to determine whether the use of the EPT leads to the collection of reliable personality data. To this end, the EPT was used to obtain two personality assessments of a sample of horses from a range of familiar raters, taken at a six months interval. This dataset was then used to evaluate the internal consistency, inter-rater reliability and test-retest reliability of the five subscales of the EPT (Gosling and Vazire, 2002; Dingemanse and Wright, 2020). This was done in order to inform subsequent chapters of this work and ensure physiological correlates were compared against valid and reliable personality scores.

The second aim of this thesis, addressed in Chapter 4, was to investigate autonomic and HPA axis reactivity to stressors as a potential correlate of personality. To this end, heart rate, heart rate variability and salivary cortisol responses were recorded in the same sample of horses during experimental exposure to mild stressors, in order to evaluate individual autonomic and HPA axis reactivity. Physiological reactivity to the stressors was then compared to EQP scores on the Agreeableness, Neuroticism and Extraversion scales. In addition to its contribution towards the overarching aim of the thesis, this chapter was also expected to determine whether the EPT has predictive validity for physiological reactivity to stressors, in addition to its predictive validity for behavioural reactivity (Ijichi et al., 2013). This is an important outcome in terms of welfare if the EPT is used as a selection tool to match horses to roles.

Chronic HPA axis activity was identified as a potentially relevant correlate as basal cortisol levels have been linked with personality traits in humans (Ormel et al., 2013; Honk et al., 2003), while the coping style framework describes differences in baseline HPA axis activity between proactive and reactive copers (Koolhaas et al., 2010). Therefore, the third aim of this thesis, addressed in Chapter 5, was to explore whether baseline activity of the HPA axis system was a correlate of personality as measured by the EPT. Hair cortisol concentration was assayed from mane hair as an indicator of baseline HPA activity. A model was then used to explore potential links between basal cortisol levels and personality, while accounting for potential confounding factors known to affect hair cortisol.

Finally, dopamine levels have been identified as a neurochemical driver of human personality (DeYoung and Gray, 2009) and of coping styles in rodents (de Boer, Buwalda and Koolhaas, 2017). In addition, genetic studies have documented links between genes related to dopamine functioning and equine personality (Momozawa et al., 2005b; Ninomiya et al., 2013). Therefore, the final aim of this thesis, addressed in Chapter 6, was to investigate whether dopamine function could also be identified as a correlate of personality in the horse. To do this, spontaneous eye blink rate was measured at rest in the stable as an indicator of tonic striatal dopamine activity. A model was then used to explore potential links between spontaneous blink rate and personality as measured by the EPT, while accounting for potential confounding factors affecting blink rate.

Chapter 2 Study Population

The sample used throughout this thesis consisted of horses recruited from the Nottingham Trent University research and teaching herd. The choice of using a sample selected from the University’s herd, rather than a privately recruited sample, was guided by a number of practical and scientific considerations outlined below.

Using University owned horses made it possible to use the same sample of horses for all the studies presented in this thesis, due to the very low turnover rate of Brackenhurst Equestrian Centre compared to most private yards. Using a consistent sample was considered important to ensure the different studies presented in this thesis were fully comparable. Using the same sample of horses throughout also made it possible to bring together the results of the different studies into a coherent whole, aiming to draw a preliminary physiological profile of different personality types.

All horses at Brackenhurst Equestrian Centre are kept on the same premises and under similar management (see 2.2 Horse management below). Compared to a private yard, horses are also exposed to very few sources of variation in their day-to-day management (e.g. no travel to competitions). This was considered important as environmental factors such as management regimes are known to influence some personality traits (Hausberger et al., 2004; Lesimple et al., 2011), as well as some of the physiological parameters examined in this thesis (e.g. hair cortisol: Gardela et al., 2020)). Therefore, using a sample of horses all kept under similar management enabled us to control this potential confounding factor. Using horses from the University’s herd, rather than another large-scale private yard, also presented the additional advantage that all details of management were available in a fully transparent manner and could therefore be used to interpret results.

Brackenhurst Equestrian Centre uses an evidence-based approach to equine management; the results of welfare studies conducted on site are incorporated in practice in order to ensure high level of welfare. For instance, indoor housing is designed to allow visual, olfactory and tactile contact with conspecifics for all horses kept in individual stables, as well as paired housing where appropriate, to avoid the negative impact of single housing with no contact (Yarnell et al., 2015). High levels of welfare were considered important for the sample as changes in physiological reactivity to stressors and baseline activity of the HPA axis are documented in response to poor welfare conditions and exposure to chronic stress (Pawluski et al., 2017; Mormède et al., 2007). While the interaction of personality, welfare conditions and physiology is a valid and important research question, it was considered to be beyond the scope of this preliminary investigation into the relationship between personality and physiological characteristics. Therefore, the known high welfare standard of the Brackenhurst Equestrian Centre herd was considered advantageous for this investigation in order to control for the potential confounding impact of welfare status.

The Brackenhurst Equestrian Centre herd is managed day to day by an experienced group of technical staff with equine management qualifications and is used for teaching purposes by academics who specialise in Equine Science. Therefore, using the Brackenhurst Equestrian Centre herd gave us access to a comparatively large number of horses whose personality could be assessed by the same, large group of highly qualified caregivers. This was deemed important in order to obtain a reliable measure of the horses' personality. Indeed, the quality of subjective personality assessment is known to depend on the rater's familiarity not only with the target animal, but also with the target species (Funder, 1995).

Sample size calculations were carried out in SPSS v.27 (SPSS Inc, Chicago, IL) to contextualise sample selection and inform interpretation of results in terms of statistical power. Papers targeted for inclusion in this analysis were those that investigated correlations between equine personality factors as measured through trait-rating and potential physiological correlates included in this thesis. Momozawa *et al.* (2003) report a correlation of 0.318 between the heart rate response to the presentation of a novel stimulus and personality factor “Anxiety”, extracted from their own trait-based equine personality questionnaire. In addition, Sauveroche *et al.* (2020) reported correlations of -0.31, -0.34 and -0.46 between hair cortisol concentration in the mane and three personality factors as measured by Lloyd *et al.*’s Horse Personality Questionnaire (2008). Finally, Roberts *et al.* (2016) reported correlations of 0.202 and 0.215 between spontaneous eye blink rate and two personality factors extracted from their own trait-based equine personality questionnaire. Overall, the strength of correlation between physiological parameters and personality factors therefore ranged between 0.2 and 0.46, with most correlations being weak to moderate ($r \approx 0.3$). Sample size calculations revealed that an approximate sample size of $n=35$ ($r=0.46$) to $n=190$ ($r=0.2$) would be needed to detect these effects with a power of 80% and a significance level of 0.05. A weak to moderate correlation with $r \approx 0.3$ could be detected with a power of 80% and a significance level of 0.05 using an approximate sample size of $n \approx 80$. This sample size analysis therefore suggests that the current study is likely underpowered to detect the effects investigated. However, it should be noted that none of the papers referenced in this analysis used the EPT, and that the equine personality questionnaires used instead all lacked elements of validation (reviewed in section 1.4.2). It is therefore unclear whether their conclusions are dicrectly comparable to the current study.

2.1 SAMPLE DEMOGRAPHICS

The sample used throughout this thesis consisted of 25 horses. Although this was a convenience sample that may not be generalizable to all sectors of the equine industry, demographic characteristics were comparable with those of the leisure/riding school horse population in the UK (Hockenhull and Creighton, 2013; Hotchkiss, Reid and Christley, 2007). It consisted of 15 geldings and 8 mares, with no stallions. All horses were adults, aged 4 to 21 years old at the time of inclusion in the sample (mean age in June 2018: 12.0 ± 4.1 years). Fourteen breeds were represented (Table 2.1), including Irish Sports Horse (n=5), Connemara (n=4), Cob (n=3), British Warmblood (n=2) and Warmblood cross (n=1), Thoroughbred (n=2) and Thoroughbred cross (n=1). All horses were fully trained for handling and riding.

All horses were part of Nottingham Trent University's research and teaching herd and were kept at Brackenhurst Equestrian Centre. At the time of first inclusion in the sample in June 2018, the horses had been housed at Brackenhurst Equestrian Centre for 3.0 ± 2.0 years on average (Table 2.1). At that time, the most recent arrivals had been at Brackenhurst Equestrian Centre for a minimum of 2 months and were deemed fully settled in the facilities and yard routine by the yard manager. Measures that may be affected by relocation stress (hair cortisol: Gardela et al., 2020) or the familiarity of staff with horses in the sample (personality: Funder, 1995) were collected after a longer delay, in September 2019 and April 2020 respectively (Table 2.1; Table 2.2). The horses' full history prior to arrival at Brackenhurst Equestrian Centre was known for some but not all horses in the sample. The majority of the horses in the sample had previously been privately owned and used for leisure riding or amateur competition. Most horses (n=22) had been started under saddle prior to their arrival at Brackenhurst Equestrian Centre and details of the methods used for their initial training were not known.

There were minimal variations in the sub-samples used for the different phases of data collection (Table 2.2). For some studies, some horses had to be excluded for ethical reasons. For instance, horses known to be highly reactive to clipping were removed from the sample for the sham clipping study to avoid distressing them. In addition, 2 horses were used to pilot the protocols used for some studies and were therefore removed from the final analysed sample for these studies. Finally, horses were removed from the sample if they left the care of Brackenhurst Equestrian Centre, even in cases when they subsequently returned or remained locally available. This was to ensure all horses in the sample were kept under consistent management. One horse was therefore removed from the sample after a period out at stud (n=1, Table 2.1). Two additional horses were removed from the sample after being sold on (n=1) and euthanised following an injury (n=1).

2.2 HORSE MANAGEMENT

2.2.1 HOUSING AND NUTRITION

Horses were kept in a mixture of indoor and outdoor housing, in different proportions at different stages of the academic year (see 2.2.2 Routines below).

2.2.1.1 Indoor housing and nutrition

When kept indoors, most horses in the sample (n=20) were kept in individual stables (3.5*4.1m: n = 15; 3.6*3.7m: n = 5) in American barns. All stables allowed visual and auditory contact with conspecifics, as well as tactile contact through barred windows or half-walls. One horse was kept individually in a crew yard (7.3*9.2m) in an American barn, allowing the same level of contact with conspecifics. Finally, 4 horses were kept in pairs in “combi barns”, consisting of a small outdoor paddock with an all-weather surface and a field shelter (3.6*7.2m). All stables, crew yards and field shelters were bedded with dust extracted wood shavings and were cleaned of urine and droppings twice a day. Horses were allocated to a particular stable or type of housing based on their individual needs and informally assessed preferences and were moved if Brackenhurst Equestrian Centre staff felt they were negatively affected by their surroundings.

Ad lib water was available at all times and horses were fed a forage-based diet following National Research Council guidelines (2007). The main source of forage was hay or haylage, fed at 2% of horses' body weight (National Research Council, 2007). The daily forage ration was split into 2 to 3 meals a day and fed in haynets (n=24) or a slow feeder (Haygain forager, n=1). All horses received forage at 7am and 4pm. When horses were kept indoors overnight, a third forage meal was given at 9pm.

Horses also received a hard feed used to balance the forage diet. In the academic years 2017-2018 and 2018-2019 horses were fed twice daily. The morning feed was delivered at 7am and consisted of pasture cubes, while the evening feed was delivered at 5pm and consisted of Ovi chaff, pasture cubes and sugar beets, to individual requirements. The feeding regime was changed for all horses ahead of the academic year 2019-2020. This was because the previous model exceeded the horses' energy requirements, leading to weight gain. From 2019-2020 onward, hard feed was delivered once daily in the morning, and consisted of chaff and a balancer (TopSpec Lite Feed Balancer) for all horses, supplemented with pasture cubes only for those horses with higher energy requirements.

2.2.1.2 Outdoor housing and nutrition

When kept outdoors, horses were turned out on grass pastures. All horses were turned out in stable single-sex groups of 2 (n = 11), 3 (n = 6) or 4 (n = 8). Group composition was determined based on individual compatibility, as assessed informally by the Brackenhurst Equestrian Centre staff. The pasture for one group of 4 horses was managed as an Equicentral set up. This management system provides horses with free choice access to a grass paddock as well as a dry lot on which shelter, hay and *ad lib* water are available (Myers and Myers, 2021). All other groups were kept on traditionally managed grass pastures, with field shelters and *ad lib* water troughs. The main source of forage was pasture grass. If grass ran low, horses were supplemented with *ad lib* hay. Hay was presented in a Monster Hay Feeder to keep the bale dry and wrapped in a slow feeder haynet (Trickle Net) to regulate ingestion and minimise wastage.

2.2.2 **ROUTINES**

Horse management at Brakenhurst Equestrian Centre was split into 3 distinct routines during different stages of the academic year, with the horses experiencing different levels of work and turnout. Details of all 3 routines are provided below. In subsequent chapters, the routine applicable during data collection will be identified so the reader can refer back to this section for more detail.

2.2.2.1 Academic year

Horses were kept in the routine referred to as the academic year routine from late September to the end of May, coinciding with the start and end of classes for Further Education students. During the academic year, horses were kept indoors during the day (7am-5pm), in the indoor facilities described in Section 2.2.1.1. Whenever the weather permitted, they were turned out on the grass pastures at night 7 days a week; outdoor turn out was as described in Section 2.2.1.2. Overnight turnout was restricted in periods of very wet weather to preserve pasture ground. During these periods, horses were kept in their indoor housing overnight. Horses were fed a forage-based diet supplemented with a balancer a described in Section 2.2.1.1.

During the academic year period horses were in work as riding school horses. Their main role was to be used in equine management and riding lessons for both Further Education and Undergraduate degree students. Horses were exercised on average 2 hours per day on week days, with a maximum workload of 3 hours per day per horse. Exercise included ridden work such as flatwork and jumping, as well as ground work (e.g. lunging, long lining, natural horsemanship exercises). Records were kept to ensure horses were assigned a balanced workload, taking into account individual differences in fitness and ability. On days when horses were not needed for lessons, they were exercised on a horse walker or turned out in a paddock instead.

2.2.2.2 Summer break

From the end of May to 1st August horses were kept in the summer break routine. During this period horses were given time off work, to coincide with the student's summer break. Horses were kept outdoors on the grass pastures described in Section 2.2.1.2 at all times, and only brought on to the yard on an *ad hoc* basis. During the summer break, horses were not exercised. However, free movement could be expressed at all times on the pastures.

2.2.2.3 Summer conditioning programme

The summer conditioning programme took place from 1st August to the end of September. During this period, horses were brought back into work by the Brackenhurst Equestrian Centre staff, ahead of the start of the academic year. Horses were kept in their indoor housing during the day (7am-5pm) on weekdays (Monday - Friday). They were turned out on their grass pastures in their stable groups at night and at the weekend. When stabled on weekdays, the nutrition regime used during the academic year was followed. When horses were out at the weekend, they were fed on pasture grass and supplemented with *ad lib* hay if grass ran low.

During the summer conditioning programme, horses followed a programme of daily work designed to build their fitness back up ahead of the start of the academic year. All exercise was given by staff members of Brackenhurst Equestrian Centre. The programme started with 40 minutes work sessions and progressively built up to 1-hour sessions. Exercise type was a mixed rota of horse walker, lunge sessions and ridden work, with minor adjustments to account for individual limitations (e.g. some horses could not be lunged due to previous injuries and were ridden instead). Exercise intensity was built up gradually, with the full sessions conducted in walk at the start of the program and aiming to have horses ready to canter by the end of September.

*Table 2.1 – Demographic details of the horses in the sample. Sex is indicated as M for a mare and G for a gelding; there were no stallions in the sample. Horses were kept at Brackenhurst Equestrian Centre continuously from their date of arrival. * identifies horses who left Brackenhurst Equestrian Centre during the period of data collection; those horses were excluded from further data collection phases following their departure, even if they subsequently re-joined the Centre.*

	Breed	**Sex**	**Year of birth**	**Arrived at Brackenhurst**
Princess	Welsh partbred	M	2010	2017
Jubilee	Cob	M	2010	2017
Greta	Haflinger	M	2006	2013
Ruby	Sufolk Punch	M	2014	2016*
Bannagh	Irish Sports Horse	G	2008	2015
Henry	Thoroughbred x Irish Sport Horse	G	2004	2017
Gunner	Irish Sports Horse	G	2004	2016*
Jordan	Appaloosa	G	2000	2017
Tom	Connemara	G	2005	2015
Louie	Sports pony	G	2010	2018
Penny	Thoroughbred	M	1997	2013
Barbie	Trakehner x Warmblood	M	2008	2011
Monty	Irish Sports Horse	G	2001	2016
Jessie	Hanoverian	M	2013	2015
Jasmine	Cob	M	2008	2013
Hercules	Irish Sports Horse	G	2007	2016
Joel	Cob	G	2010	2018*
Woody	British Warmblood	G	2005	2017

Bracken	Connemara	G	2004	2014
Rocky	Connemara	G	2006	2015
Bobby	Thoroughbred x Warmblood	G	2007	2013
Daniel	Irish Sports Horse	G	2005	2013
Zazou	British Warmblood	G	2004	2013
Lily	Thoroughbred	M	2000	2012
Misty	Connemara	G	2003	2016

Table 2.2 – List of data collection phases for the thesis. In the top section, the date of data collection took place is indicated, as well as the the sample size used. In the bottom section, horses who were included in the sample are marked by a tick. Horses who are not marked by a tick were excluded from the sample for the corresponding phase of data collection; the reason for exclusion is listed instead.

	Personality Questionnaire (Chapters 3-6)	**Novel Object test** (Chapter 4)	**Startle test** (Chapter 4)	**Sham Clipping** (Chapter 4)	**Hair cortisol** (Chapter 5)	**Spontaneous blink rate** (Chapter 6)
Date	04-10/2020	06/2018	06/2018	10-12/2018	09/2019	08/2019
Sample size	n = 25	n = 23	n = 23	n = 21	n = 24	n = 20
Princess	✓	✓	✓	✓	✓	✓
Jubilee	✓	✓	✓	✓	✓	✓
Greta	✓	✓	✓	✓	✓	✓
Ruby	✓	✓	✓	✓	Pregnant	Pregnant
Bannagh	✓	✓	✓	Ethics: overly reactive to clipping	✓	✓
Henry	✓	✓	✓	✓	✓	✓
Gunner	✓	✓	✓	Injury	✓	Deceased
Jordan	✓	✓	✓	✓	✓	✓
Tom	✓	✓	✓	✓	✓	✓
Louie	✓	✓	✓	✓	✓	✓
Penny	✓	✓	✓	Ethics: overly reactive to clipping	✓	✓

Barbie	✓	✓	✓	✓	✓	✓
Monty	✓	✓	✓	✓	✓	✓
Jessie	✓	✓	✓	✓	✓	✓
Jasmine	✓	✓	✓	✓	✓	✓
Hercules	✓	✓	✓	✓	✓	✓
Joel	✓	✓	✓	✓	✓	Sold on
Woody	✓	✓	✓	✓	✓	✓
Bracken	✓	✓	✓	✓	✓	✓
Rocky	✓	✓	✓	✓	✓	✓
Bobby	✓	✓	✓	✓	✓	✓
Daniel	✓	✓	✓	✓	✓	✓
Zazou	✓	✓	✓	✓	✓	✓
Lily	✓	Used in pilot	Used in pilot	Used in pilot	✓	Used in pilot
Misty	✓	Used in pilot	Used in pilot	Used in pilot	✓	Used in pilot

Chapter 3 Internal consistency and reliability of the Equine Personality Test

3.1 INTRODUCTION

Identifying correlates of personality is only meaningful and of applied importance if the personality assessment tool provides a truly valid and reliable measure of target animals' underlying behavioural tendencies. Therefore, before correlates of personality can be investigated, it is crucial to establish whether the personality assessment tool used satisfies the four validity and reliability criteria emerging from psychometric research (Gosling and Vazire, 2002; Dingemanse and Wright, 2020). When using a trait-based subjective assessment, these criteria are: (1) each subscale measuring a personality factor should have internal consistency, i.e. measure a single underlying construct (Simms and Watson, 2007); (2) each subscale should have concurrent validity, i.e. scores on the scale should reflect the expression of conceptually related behaviour (Gosling and Vazire, 2002); (3) reliable items should be used, so that independent raters agree in their evaluation of a familiar target animal (Gosling and Vazire, 2002); and (4) scores for a target animal should be consistent over time (Dingemanse and Wright, 2020) (see Section 1.2.3 for more detail). To date, no published equine personality questionnaire has been evaluated on all four criteria of validity and reliability outlined here. In particular, the equine personality assessment tool chosen for use in this work, the Equine Personality Test (EPT; Ijichi et al., 2013) has only been evaluated against one of these critera in a peer reviewed publication.

The concurrent and predictive validity of the Neuroticism and Extraversion scales of the EPT has been demonstrated (Ijichi et al., 2013; Ijichi, Collins and Elwood, 2014). This is the criterion equine personality questionnaires are most checked against, with personality scores compared with behaviour in standardised tests (Anderson et al., 1999; Seaman et al., 2002; Momozawa et al., 2003; Visser et al., 2003) or naturally occurring behaviour (Morris et al., 2002; Lloyd et al., 2007). In the case of the EPT, Neuroticism and Extraversion scores predict conceptually-related behaviour in standardised tests including a novel object test, a startle test and a novel handling test (Ijichi et al., 2013). Neuroticism predicts the intensity of the startle response and tends to predict the time to complete the novel handling test (Ijichi et al., 2013). By contrast, Extraversion predicts whether an active or passive strategy is adopted when expressing refusal in a handling test, and is hypothesised to reflect coping style in the horse (Ijichi et al., 2013). These two scales also predict the expression of horses' pain responses in a veterinary context (Ijichi et al., 2014). These scales of the EPT therefore satisfy the second criterion set by Gosling & Vazire (2002) for the validity of psychometric tools used to measure animal personality, and compare favourably with other equine personality questionnaire available in the range of behaviours they have been related to.

The internal consistency, inter-rater reliability and test-retest reliability of the EPT have not yet been evaluated. Despite the importance of demonstrating that the traits measured by personality assessments are stable over time (Dingemanse and Wright, 2020), to date no equine personality questionnaires has been checked for test-retest reliability. However, internal consistency on par with thate of human Five Factor Model questionnaires (e.g. Mccrae and Costa, 1987; Scandell, 2000) is reported for two equine personality questionnaires (Momozawa et al., 2005a; McGrogan, Hutchison and King, 2008). The wide range of statistical indices used to assess the inter-rater reliability of equine personality questionnaires makes comparisons between questionnaires challenging (e.g. Anderson et al., 1999; Morris et al., 2002; Lloyd et al., 2007). However, Lloyd et al. (2007) and Morris et al. (2002) both report good levels of inter-rater agreement, with 72.1% of horses being agreed on by the raters and between-subjects correlation coefficients higher than 0.37, respectively. In addition, Anderson et al. (1999) report poor inter-rater agreement for their questionnaire adapted from the Donkey Temperament Survey (French, 1993). However, they use a much more stringent threshold for acceptable agreement (Anderson et al., 1999). Overall, this suggests that good inter-rater reliability can be achieved when using equine personality questionnaires. This is coherent with results in other species that suggest animal personality assessment tools can achieve comparable inter-rater reliability to human questionnaires (Gosling, 2001). In line with findings in other species (Gosling, 2001), the degree of familiarity of the rater with the target horse, as well as the variety of contexts in which the individual could be observed, may also impact on the inter-rater reliability of scoring (Lloyd et al., 2007). Therefore, although the EPT has not yet been evaluated against the criteria of internal consistency, inter-rater reliability and test-retest reliability, benchmarks are available in the literature to compare its performance on these criteria against that of other published equine personality questionnaires.

While the EPT has been shown to produce valid measures of personality, little is known about its internal consistency and reliability. The aim of this chapter was therefore to investigate how well the EPT meets the criteria for internal consistency and reliability for a personality assessment tool. Personality data of 25 horses was collected from 6 raters using the EPT and used to compute indices of internal consistency, inter-rater reliability and test-retest. The EPT has been developed using a psychometric approach (Ijichi et al., 2013; Creighton, personal communication); therefore, it was hypothesised that all subscales would show good internal consistency, inter-rater reliability and test-retest reliability, on par with those documented for other animal personality assessment tools. The findings of this chapter aimed to inform the rest of the current study by ensuring the personality assessment tool chosen for use in this project could generate valid and reliable personality data.

3.2 METHODS

3.2.1 SUBJECTS

Personality data were collected for all horses used in this project (n=25). There were 9 mares and 16 geldings; mean age was 14.0 ± 4.1 years. 10 breeds were represented, including Irish Sports Horse (n=5), Connemara (n=4), Cob (n=3), British Warmblood (n= 2) and Thoroughbred (n=2). This represents a good cross-section of the British equine population (Hockenhull and Creighton, 2013; Hotchkiss, Reid and Christley, 2007). All horses were recruited from Brackenhurst Equestrian Centre. During the academic year these horses are used for equine management and equitation teaching in a riding school-like setting. All horses lived on the same premises and were kept under the same management regime. For more demographic and management details about the horses please refer to section 2.1; the rationale behind the choice of the study population for the thesis is explained in further detail in the introduction of Chapter 2.

3.2.2 RATERS

Six raters were recruited via email from a pool of Nottingham Trent University staff meeting two inclusion criteria aimed at maximising the accuracy of the personality assessment. First, raters had to be familiar with the horses in the sample. All had known the horses for a minimum of two years and interacted with them on a daily to weekly basis. In addition, raters recruited had to possess a strong knowledge of the species as a whole (Gosling, 2001). All raters had 10+ years of professional experience in the equine industry in a wide variety of roles (instruction/coaching, competition, training, and day-to-day management). Three raters, ES, CH and JBF, were recruited from the Equine Technical team; ES later indicated that she had completed some of the questionnaires with input from a 6th rater, AH, also from the Equine Technical team. The Equine Technical team is responsible for the day-to-day care of the horses; ES, CH and JBF were therefore primary caregivers for all horses in the sample. The 3 other raters were members of the academic staff in Further Education (LT) and Higher Education (TC and SH). All three regularly used the horses in the sample to teach management and riding lessons. It should be noted that the recruitment process differed from the one described in Ijichi *et al.* (2013, 2014) as the current study aimed to gather personality assessments of each horse in the sample from a range of raters, rather than a single rater.

Raters were contacted via email to invite them to take part in the study and all accepted. They were then sent a link to a OneDrive folder containing a digital version of the questionnaire ready for each horse. They were provided with technical guidance to access and edit the online files and could obtain further explanations on how to fill in the questionnaire from the experimenter upon request. Raters were instructed to only fill in the questionnaire for horses they felt confident they were familiar with. As a result, 3 of the raters elected not to carry out the personality assessment for some of the horses. They were also invited to leave informal feedback either via email or within the questionnaire forms if they felt unsure of any of their ratings for some horses or questions. While this feedback was not used in calculating individual horse's personality scores, it was referred to when interpreting results. All raters completed the questionnaire independently from each other and did not discuss their assessments of the horses between themselves. While available to answer technical questions, the experimenter did not provide any input into the personality assessment.

3.2.3 DATA COLLECTION CALENDAR FOR INTER-RATER AND TEST-RETEST RELIABILITY STUDIES

ES (with AH's help) completed the questionnaires for the 25 horses in the sample in August 2019. CH, JBF, TC, LT and SH all completed the questionnaires a first time for the inter-rater reliability study in April 2020. Subsequently, CH and JBF completed the questionnaires again 6 months later, in October 2020, for the test-retest reliability study. ES was not asked to take part in the test-retest study as the extent of AH's help in filling the first batch of questionnaires was unknown and could not be reproduced. The three instructors (LT, TC and SH) were not asked to take part in the test-retest reliability study either. This was because they had not been in regular contact with the horses in the sample between April and October 2020, due to the impact of the COVID19 lockdown (March to June) followed by the University summer break (June to October).

3.2.4 QUESTIONNAIRE STRUCTURE AND DATA PROCESSING

The questionnaire comprised of 22 items, divided into two sections. In the first section, raters described the target individual on a visual analogue scale between pairs of opposite adjectives (e.g. Spirited/Steady). In the second section, 5-point Likert scales were used to answer 7 questions. Precautions in accordance with psychometrics research guidelines were taken in designing the questionnaire in order to reduce potential biases from superficial scoring (Ijichi et al., 2013). Traits were presented in randomised order to avoid grouping them by factor (Ijichi et al., 2013). In addition, the polarity of pairs of adjectives was randomly reversed in order to avoid systematic scoring down the positive or negative end of the scale (Ijichi et al., 2013).

Once they had been filled in by the raters, online questionnaires were downloaded by the experimenter. Personality scores were calculated for each horse using the scoring protocol described in Ijichi et al., (2013) (Table 3.1). The visual analogue scales in the first section of the questionnaire were divided into 5 sections of equal length, numbered from 1 to 5 left to right. For each item, the number of the section containing the tick the rater had placed on the original visual analogue scale was recorded. For the items marked 'Reversed' in Table 3.1, the final score for the item was obtained by subtracting this number from 6. The scores given for the 7 questions in the second section of the questionnaire were also recorded. Ijichi et al., (2013) recommends that scores for each personality factor should be calculated by adding the scores of all questionnaire items associated with that factor. However, in some questionnaires (n=10) raters had overlooked one or more questions, resulting in artificially reduced factor scores for the horses concerned. Because the EPT has yet not been shown to produce consistent responses over time (test-retest reliability), it was deemed inappropriate to collect this missing data in a second sitting. In addition, discarding the questionnaires altogether would have led to an important reduction in sample size for inter-rater and test-retest reliability studies (N=11 horses) as the statistic used does not tolerate missing data. Therefore, the decision was made to obtain factor scores by averaging all item scores available for the factor, rather than adding them as recommended by Ijichi et al. (2013). This seemed justified as visual examination of the data suggested factors showed good homogeneity; it was therefore considered unlikely that the missing item score would be very different from those that had been provided. This was subsequently confirmed formally by the data analysis. Due to the change in scoring method implemented, continuous scores between 1 and 5 were therefore obtained for all 5 personality factors.

Table 3.1. Scoring grid for the Equine Personality Test: items associated with each personality factor. Items marked as 'Reversed' are scored by subtracting the score given to the horses from 6. Adapted from Ijichi et al. (2013).

Personality dimension	Item	Adjectives / Question	Reversed
Agreeableness	A1	Easy-going / Intolerant	✓
	A2	Argumentative / Well-mannered	
	A4	Obedient / Wayward	✓
	A6	Willing / Stubborn	✓
	A8	Gentle / Rough	✓
Neuroticism	A3	Anxious / Confident	✓
	A11	Nervous / Calm	✓
	A13	Relaxed / Tense	
	A14	Quiet / Restless	
	Q5	In general how fearful is this horse around other horses?	
Extraversion	A5	Sluggish / Forward-going	
	A7	Placid / Active	
	A9	Adventurous / Habitual	✓
	A10	Excitable / Laid-back	✓
	A12	Spirited / Steady	✓
	Q6	Generally how energetic would you say this horse is?	
Gregariousness – People	A15	Friendly / Standoffish	✓
	Q1	When it has the opportunity, how often does this horse initiate interaction with you?	
	Q2	When it has the opportunity, how often does this horse initiate interaction with other people?	
Gregariousness – Horses	Q3	When it has the opportunity, how often does this horse initiate interaction with other horses?	
	Q4	Does this horse ever show affection towards other horses?	
	Q7	Generally how dependable would you say this horse is?	

3.2.5 STATISTICAL ANALYSIS

3.2.5.1 Internal consistency and homogeneity of each personality factor subscale

This analysis was performed in SPSS v.26 (SPSS Inc, Chicago, IL). Internal consistency was assessed using Cronbach's α (Cronbach, 1951). If a questionnaire has subscales, Cronbach's α must be applied to each subscale rather than the questionnaire as a whole (Simms and Watson, 2007). Therefore, for each rater a value of Cronbach's α was calculated for each of the 5 subscales of the EPT. For each subscale, the mean ± standard deviation of the Cronbach's α's for the 6 raters were also calculated. Cronbach's α is sensitive to the direction of coding used for Likert-like data (Field, 2009). This analysis was therefore run using the coded data, with the relevant questions reversed, rather than the raw data from the questionnaires. Resulting Cronbach's α's were compared to published thresholds for acceptable internal consistency: a coefficient α higher than 0.7 is generally regarded as indicating acceptable internal consistency in a scale (Field, 2009).

The homogeneity of each subscale was also evaluated by calculating the mean and distribution of inter-item Spearman's correlation coefficients (Simms and Watson, 2007). Values obtained for the subscales of the EPT were compared to published standards for scale homogeneity: a mean inter-item correlation between 0.15 and 0.5, with a distribution of coefficients closely clustered around the mean, indicates a homogenous scale (Simms and Watson, 2007). Conversely, significant variability in the correlation coefficients could indicate multidimensionality in the scale (Simms and Watson, 2007).

3.2.5.2 Inter-rater agreement

This analysis was carried out in SPSS v.26 (SPSS Inc, Chicago, IL). Intra-class correlation (ICC) analysis was used to evaluate inter-rater reliability. ICC is the recommended method to investigate inter-rater agreement in trait rating-based animal personality studies (Vazire et al., 2007). The overall inter-rater reliability of the set of raters was evaluated through a first ICC using pooled data from both primary caregivers and instructors. Rater ES was excluded from this analysis as they had filled in some questionnaires in consultation with a third party and their scores could not be attributed with certainty to a single rater. A total of 5 raters were therefore entered into this initial analysis (CH, JBF, TC, SH, LT: k_O=5). In addition, the inter-rater reliability of primary caregivers and instructors was also compared. To this end, separate ICC analyses were carried out using the scores given by the primary caregivers on the one hand (ES, CH and JBF: k_{PC}=3), and those given by the riding instructors on the other (TC, SH and LT: k_I=3). Mean-rating (k_O=5; k_{PC}=3; k_I=3), absolute agreement, two-way random effect models were used throughout (McGraw and Wong, 1996). Model selection was based on decision trees and guidance in Hallgren (2012) and Koo and Li (2016). A two-way model was selected because all horses had been assessed by the same raters. Random effects were chosen because the raters recruited to this study were a random set of raters selected from a wider population. The type of model was set to mean-rating rather than single-rating because analysis using personality data in subsequent chapters will be carried out using the mean value of assessments obtained from all raters. Finally, the definition used was absolute agreement rather than consistency.

Inter-rater reliability was assessed for each of the 5 subscales measuring the 5 personality dimensions, as these were the outcomes used in subsequent analysis. In addition, inter-rater reliability was also assessed for each questionnaire item separately, in order to identify if some items yielded particularly high levels of disagreement between raters. For each personality factor and questionnaire item ICC estimates and their 95% confidence interval were calculated. Interpretation in terms of inter-rater reliability for the subscale or item was carried out using the thresholds for poor (ICC<0.5), moderate (0.5<ICC<0.75), good (0.75<ICC<0.9) and excellent (ICC>0.9) agreement proposed by Koo and Li (2016).

3.2.5.3 Test-retest reliability

This analysis was carried out in R version 3.6.1 (R Core Team, 2019) using the *irr* package (Gamer, Lemon and Fellows Puspendra Singh, 2019). Intra-class correlation (ICC) analysis was used to evaluate the level of test-retest reliability of the questionnaire (Koo and Li, 2016). Test-retest reliability was assessed for each of the 5 subscales measuring the 5 personality dimensions, as these were the outcomes used in subsequent analysis. In addition, test-retest reliability was also assessed for each questionnaire item separately, in order to identify if some items showed higher inconsistency over time.

Two sets of scores, for test and retest, were obtained by averaging the scores given by CH and JBF on the questionnaires they completed in April and October, respectively. The scores for test and retest were then compared using a single-ratings, absolute agreement, two-way mixed effects model (McGraw and Wong, 1996). Model selection was guided by Koo and Li (2016)'s decision tree, after Shrout & Fleiss' (1979) recommendations: for intra-rater reliability studies a two-way model is selected because all subjects are rated by the same raters, with mixed effects as rater selection is not random. In addition, absolute agreement rather than consistency should be evaluated when investigating intra-rater reliability. Here, single ratings rather than mean ratings were used, to account for the fact that in subsequent studies the personality scores used will only result from a single administration of the EPT rather than be averaged across a number of retests.

3.3 RESULTS

3.3.1 DESCRIPTIVE STATISTICS

Figure 3.1 shows the distribution of scores given to the 25 horses, averaged across the 6 raters, for each of the 5 personality factors assessed by the Equine Personality Test. Scores were evenly distributed along the scale for Neuroticism (mean score: 2.4 ± 0.7) and Extraversion (mean score: 3.1 ± 0.7). However, most horses in the samples scored on the highest half of the scale for Agreeableness (mean score: 3.7 ± 0.8), Gregariousness towards People (mean score: 3.6 ± 0.8) and Gregariousness towards Horses (mean score: 3.4 ± 0.4). There was reasonable variation in the scores for Agreeableness, Neuroticism, Extraversion and Gregariousness towards People, with most of the range of the scales represented in the sample. However, there was much less variation in the scores for Gregariousness towards Horses.

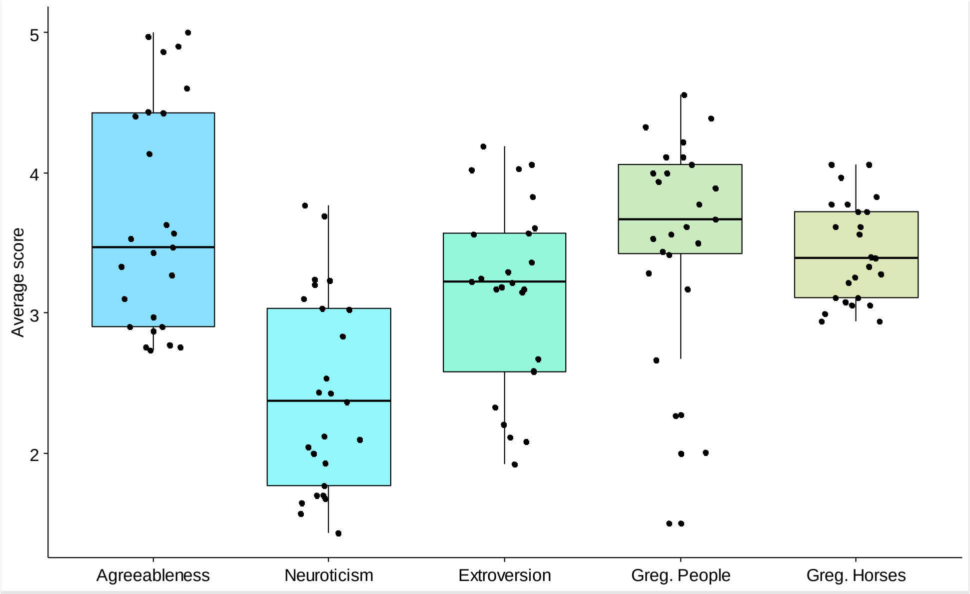

Figure 3.1 – Boxplots showing the distribution of scores for the 25 horses, averaged across the 6 raters, for the 5 personality factors assessed by the Equine Personality Test.

The three instructors (raters TC, LT and SH), who were familiar with the horses mostly in a ridden context, all expressed that they had found it challenging to score the horses on questionnaire items relating to their behaviour towards other horses (items Q3, Q4 and Q5). This concern was not shared by the primary caregivers (raters ES, CH and JBF).

3.3.2 INTERNAL CONSISTENCY AND HOMOGENEITY OF EACH PERSONALITY FACTOR SUBSCALE

Cronbach's α's for each subscale and each rater are presented in Table 3.2, along with mean and standard deviation α across the 5 raters for each subscale. Cronbach's α were relatively consistent across raters for all personality factors. The highest variation was for Gregariousness towards Horses.

Cronbach's α's were high (>0.7) for all raters for the subscales measuring Agreeableness, Neuroticism, Extraversion and Gregariousness towards People. The mean α for each of those subscales were all higher than the threshold of 0.7. However, analysis revealed that the internal consistency of some of those subscales could be improved by removing some items. Values of Cronbach's α were higher if the items "In general how fearful is this horse around other horses?" (Q5), "Adventurous/Habitual" (A9) and "Friendly/Standoffish" (A15) were removed from the Neuroticism, Extraversion and Gregariousness towards People scales, respectively.

Conversely, Cronbach's αs were very low for the scale measuring Gregariousness towards Horses, ranging from 0.12 to 0.54 with a mean of 0.39 ±0.15 across the 6 raters. For all 6 raters, removing the item Q7: "Generally how dependable would you say this horse is?" resulted in an increase of Cronbach's α above the threshold for acceptable internal consistency. Mean α across the 6 raters with Q7 removed was 0.77 ± 0.38.

Table 3.2 – Cronbach's α for each personality factor and each rater. For each factor, the lowest α between the 6 raters is presented in italics and the highest in ***bold****. For each factor the mean and standard deviation of Cronbach's α for the 6 raters are also presented.*

	Raters							
	ES	**CH**	**TC**	**LT**	**SH**	**JBF**	**Mean**	**StDev**
Agreeableness	*0.76*	0.80	0.93	0.78	0.86	**0.96**	0.85	0.08
Neuroticism	0.85	**0.89**	0.83	0.79	*0.78*	0.81	0.83	0.04
Extraversion	*0.81*	0.83	0.85	**0.92**	0.90	0.83	0.86	0.04
Greg. People	*0.84*	0.92	0.89	0.92	**0.96**	0.79	0.89	0.06
Greg. Horses	0.42	0.39	0.49	**0.54**	*0.12*	0.41	0.39	0.15

Mean inter-item correlation coefficient (± SD) was 0.64 ± 0.14 for Agreeableness, 0.61 ± 0.26 for Neuroticism, 0.56 ± 0.24 for Extraversion, 0.87 ± 0.06 for Gregariousness towards People, and 0.18 ± 0.39 for Gregariousness towards Horses. Figure 3.2 shows the mean and distribution of individual inter-item correlation coefficients for each subscale. Inter-item correlation coefficients cluster relatively closely around the mean for Agreeableness, Neuroticism, Extraversion and Gregariousness towards People. However, there is much more variability for Gregariousness towards Horses.

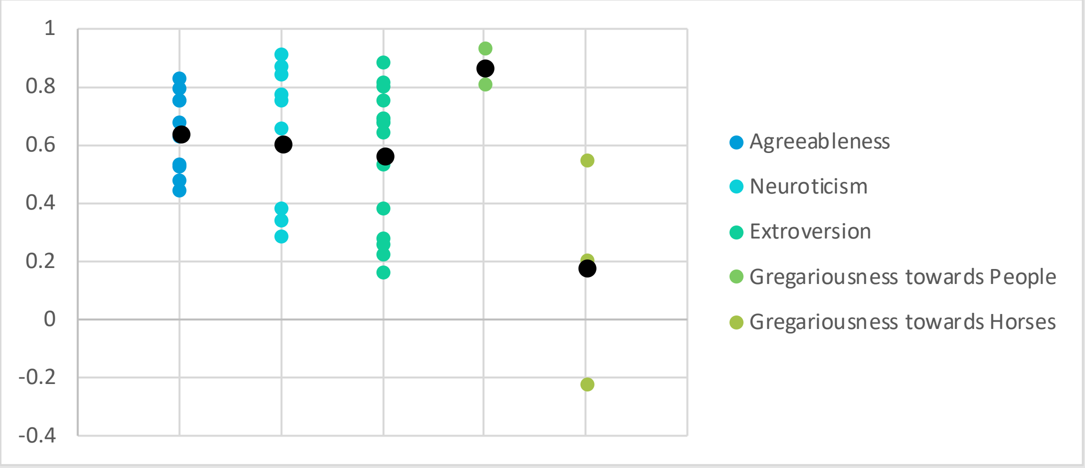

Figure 3.2 – Mean and distribution of inter-item correlation coefficients for the 5 subscales of the questionnaire. For each subscale, means are represented in black and pairwise inter-item correlation coefficients are represented in colour.

3.3.3 INTER-RATER AGREEMENT

3.3.3.1 Inter-rater agreement across the whole sample of 6 raters

ICC estimates for Agreeableness, Neuroticism, Extraversion and Gregariousness towards People were all higher than 0.75 (Table 3.3), indicating good inter-rater reliability for those four factors (Koo and Li, 2016). On the basis of the 95% confidence intervals for the ICC estimates for those 4 factors, the true level of reliability is moderate to excellent. However, the ICC estimate for Gregariousness towards Horses is lower than 0.5 (Table 3.3), and the 95% confidence interval indicates that inter-rater reliability for this factor is poor to moderate at best.

ICC estimates and their 95% confidence interval for each of the 22 questionnaire items are also presented in Table 3.3. The average ICC across all questionnaire items was 0.66 ± 0.22, ranging from 0 to 0.869. Inter-rater agreement was good for 13 items, with ICC estimates ranging from 0.754 to 0.869, and moderate for another 5, with ICC estimates ranging between 0.613 and 0.737. However, it was poor (ICC<0.5) for 4 items: one from the Neuroticism subscale (ICC_{Q5}=-0.011), one from the Extraversion subscale (ICC_{A9}=0.487), and two from the Gregariousness towards Horses subscale (ICC_{Q3}=0.346 and ICC_{Q4}=0.182). Three of these four items related to the horse's behaviour towards other horses.

3.3.3.2 Caregivers vs. instructors comparison

Comparisons between the two groups revealed that overall primary caregivers showed better inter-rater reliability than instructors (Table 3.4). At the subscale level, primary caregivers had good inter-rater agreement for 3 of the 5 personality factors (Agreeableness, Neuroticism and Gregariousness towards People), and moderate agreement for the remaining two (Extraversion and Gregariousness towards Horses). By contrast, instructors only had good agreement for two factors (Neuroticism and Extraversion), while agreement was moderate for another two (Agreeableness and Gregariousness towards People) and poor for a third (Gregariousness towards Horses). ICC coefficients were higher for primary caregivers than for instructors for all factors except Extraversion, indicating higher levels of inter-rater agreement within that group. The most obvious difference between groups was for Gregariousness towards Horses. For this factor the ICC coefficient was 0.562 for primary caregivers (moderate agreement) but only 0.391 for instructors (poor agreement).

Similarly, at the item level, good levels of inter-rater agreement were observed more often for primary caregivers than for the instructors. For primary caregivers, reliability was good for 8 questionnaire items, moderate for 12 and poor for only 2. By contrast, for instructors, reliability was good for only 2 items, while it was moderate for 15 and poor for 5. For all but 5 items, ICC coefficients were higher for primary caregivers than instructors, indicating better levels of inter-rater agreement.

Table 3.3 - Results of intra-class correlation analyses for all subscales and individual items in the questionnaire. A mean-rating (k=5), absolute-agreement, two-way random-effects model was used. For each subscale or item, the ICC estimate, 95% confidence interval, and interpretation in terms of inter-rater reliability for the item are presented. The subscript $^{?}$ highlights subscales or items for which the ICC estimate indicates poor inter-rater reliability.

			95% confidence interval		
	n	**ICC**	Lower	Upper	**Reliability**
Agreeableness	21	0.848	0.715	0.930	Moderate to excellent
Easy-going/Intolerant	21	0.869	0.754	0.94	Good to excellent
Argumentative/Well-mannered	21	0.774	0.578	0.869	Moderate to good
Obedient/Wayward	21	0.788	0.607	0.902	Moderate to excellent
Willing/Stubborn	21	0.764	0.559	0.892	Moderate to good
Gentle/Rough	21	0.613	0.29	0.82	Poor to good
Neuroticism	21	0.848	0.715	0.930	Moderate to excellent
Anxious/Confident	21	0.831	0.679	0.923	Moderate to excellent
Nervous/Calm	21	0.782	0.593	0.9	Moderate to good
Relaxed/Tense	20	0.765	0.553	0.895	Moderate to good
Quiet/Restless	19	0.771	0.552	0.9	Moderate to good
How fearful is this horse around other horses?	20	-0.011$^{?}$	-0.47	0.443	Poor
Extraversion	21	0.806	0.640	0.911	Moderate to excellent

Sluggish/Forward-going	21	0.663	0.388	0.842	Poor to good
Placid/Active	20	0.798	0.618	0.909	Moderate to excellent
Adventurous/Habitual	21	0.487[?]	0.036	0.736	Poor to moderate
Excitable/Laid-back	17	0.705	0.415	0.878	Poor to good
Spirited/Steady	20	0.757	0.544	0.89	Moderate to good
How energetic would you say this horse is?	20	0.696	0.423	0.863	Poor to good
Gregariousness towards people	21	0.829	0.681	0.921	Moderate to excellent
Friendly/Standoffish	20	0.797	0.611	0.909	Moderate to excellent
How often does this horse initiate interaction with you	21	0.754	0.539	0.887	Moderate to good
How often does this horse initiate interaction with other people	21	0.770	0.562	0.895	Moderate to good
Gregariousness towards horses	21	0.498[?]	0.140	0.752	Poor to moderate
How often does this horse initiate interaction with other horses?	21	0.346[?]	0.015	0.643	Poor to moderate
Does this horse ever show affection towards other horses?	21	0.182[?]	-0.095	0.499	Poor
How dependable would you say this horse is?	20	0.737	0.501	0.884	Moderate to good

Table 3.4 – Comparison of the inter-rater reliability of primary caregivers vs. instructors for all subscales and individual items in the questionnaire. Mean-rating (k=3), absolute-agreement, two-way random-effects models were used to carry out separate intra-class correlation analyses for the two groups. For each subscale or item the sample size, ICC estimate, and interpretation in terms of inter-rater reliability for the item are presented. Italics highlight subscales or items for which instructors had better inter-rater reliability than primary caregivers.

	Primary caregivers			**Instructors**		
	n	**ICC**	**Reliability**	**n**	**ICC**	**Reliability**
Agreeableness	25	0.849	Good	21	0.724	Moderate
Easy-going/Intolerant	*25*	*0.787*	*Good*	*21*	*0.795*	*Good*
Argumentative/Well-mannered	25	0.807	Good	21	0.584	Moderate
Obedient/Wayward	25	0.824	Good	21	0.594	Moderate
Willing/Stubborn	*25*	*0.520*	*Moderate*	*21*	*0.571*	*Moderate*
Gentle/Rough	25	0.733	Moderate	21	0.341	Poor
Neuroticism	25	0.792	Good	21	0.777	Good
Anxious/Confident	25	0.791	Good	21	0.694	Moderate
Nervous/Calm	25	0.736	Moderate	21	0.618	Moderate
Relaxed/Tense	25	0.786	Good	20	0.665	Moderate
Quiet/Restless	24	0.762	Good	19	0.72	Moderate
How fearful is this horse around other horses?	25	0.265	Poor	20	-0.032	Poor

Extraversion	*25*	*0.720*	*Moderate*	*21*	*0.766*	*Good*
Sluggish/Forward-going	24	0.614	Moderate	21	0.576	Moderate
Placid/Active	*25*	*0.591*	*Moderate*	*20*	*0.778*	*Good*
Adventurous/Habitual	25	0.541	Moderate	21	0.435	Poor
Excitable/Laid-back	21	0.731	Moderate	21	0.63	Moderate
Spirited/Steady	24	0.718	Moderate	21	0.638	Moderate
How energetic would you say this horse is?	*25*	*0.518*	*Moderate*	*20*	*0.653*	*Moderate*
Gregariousness towards people	25	0.827	Good	21	0.708	Moderate
Friendly/Standoffish	25	0.831	Good	20	0.604	Moderate
How often does this horse initiate interaction with you	25	0.687	Moderate	21	0.675	Moderate
How often does this horse initiate interaction with other people	25	0.754	Good	21	0.628	Moderate
Gregariousness towards horses	25	0.562	Moderate	21	0.391	Poor
How often does this horse initiate interaction with other horses?	25	0.511	Moderate	21	0.293	Poor
Does this horse ever show affection towards other horses?	25	0.494	Poor	21	0.147	Poor
How dependable would you say this horse is?	*24*	*0.571*	*Moderate*	*20*	*0.665*	*Moderate*

3.3.4 TEST-RETEST RELIABILITY

ICC estimates and 95% confidence intervals for the test-retest reliability of the 5 subscales and 22 questionnaire items are presented in Table 3.5. The ICC estimates the subscales measuring Neuroticism, Extraversion and Gregariousness towards People were all greater than 0.9 (Table 3.5), and the 95% confidence intervals indicated that test-retest reliability for these subscales was good to excellent. However, the ICC estimates for the subscales measuring Agreeableness and Gregariousness towards Horses were greater than 0.75 but lower than 0.9. The 95% confidence intervals for these subscales indicated moderate to excellent test-retest reliability for Agreeableness but only moderate to good reliability for Gregariousness towards Horses.

At the items level, 14 items showed good or excellent test-retest reliability ($ICC>0.75$). A further 8 performed more poorly, with 7 showing moderate reliability ($0.5<ICC<0.75$) and one showing poor reliability ($ICC_{Q5}<0.5$). The Agreeableness, Neuroticism and Extraversion subscales all had a minority of poorly performing items (up to 2 per subscale). However, all items on the Gregariousness towards Horses subscale performed poorly, with ICC estimates lower than 0.75 (moderate reliability) and 95% confidence intervals indicating poor to good reliability.

Table 3.5 - Results of test-retest reliability analyses for all subscales and individual items in the questionnaire. A single-rating, absolute agreement, two-way mixed effects model was used to carry out intra-class correlations analyses. For each subscale or item the ICC estimate, 95% confidence interval, and interpretation in terms of test-retest reliability for the subscale or item are presented. The subscript ? highlights subscales or items for which the ICC estimate indicates poor test-retest reliability.

			95% confidence interval		
	n	**ICC**	Lower	Upper	**Reliability**
Agreeableness	25	0.868	0.699	0.942	Good
Easy-going/Intolerant	25	0.821	0.639	0.917	Good
Argumentative/Well-mannered	25	0.717	0.283	0.884	Moderate
Obedient/Wayward	25	0.777	0.525	0.899	Good
Willing/Stubborn	25	0.700	0.430	0.855	Moderate
Gentle/Rough	25	0.759	0.526	0.886	Good
Neuroticism	25	0.903	0.792	0.956	Excellent
Anxious/Confident	25	0.915	0.816	0.962	Excellent
Nervous/Calm	25	0.846	0.682	0.929	Good
Relaxed/Tense	25	0.786	0.574	0.900	Good
Quiet/Restless	25	0.647	0.342	0.828	Moderate
How fearful is this horse around other horses?	25	$0.352^{?}$	-0.047	0.653	Poor

Extraversion	25	0.910	0.808	0.959	Excellent
Sluggish/Forward-going	25	0.870	0.731	0.940	Good
Placid/Active	25	0.815	0.626	0.914	Good
Adventurous/Habitual	25	0.858	0.704	0.935	Good
Excitable/Laid-back	25	0.844	0.677	0.928	Good
Spirited/Steady	25	0.840	0.659	0.927	Good
How energetic would you say this horse is?	25	0.724	0.469	0.868	Moderate
Gregariousness towards people	25	0.922	0.818	0.966	Excellent
Friendly/Standoffish	25	0.890	0.766	0.950	Good
How often does this horse initiate interaction with you	25	0.777	0.549	0.896	Good
How often does this horse initiate interaction with other people	25	0.820	0.575	0.923	Good
Gregariousness towards horses	25	0.784	0.572	0.898	Good
How often does this horse initiate interaction with other horses?	25	0.592	0.217	0.806	Moderate
Does this horse ever show affection towards other horses?	25	0.703	0.440	0.856	Moderate
How dependable would you say this horse is?	25	0.722	0.469	0.866	Moderate

3.4 DISCUSSION

The Equine Personality Test has previously been shown to have good predictive reliability (Ijichi et al., 2013). However, further checks on its internal consistency, inter-rater reliability, and test-retest reliability had not yet been carried out. The aim of this study was to evaluate the EPT's performance on these three criteria for the sample of horses and raters used in this thesis. To this end, 6 raters were asked to use the EPT to assess 25 horses, with 2 raters carrying out the assessment twice over a period of 6 months. Cronbach's α and intra-class correlations analyses were used to analyse scale internal consistency and inter-rater and test-retest reliability, respectively. While the Agreeableness, Neuroticism, Extraversion and Gregariousness towards People subscales performed well on all three criteria, the Gregariousness towards Horses subscale failed to meet the criteria.

The Agreeableness, Neuroticism, Extraversion and Gregariousness towards People subscales had Cronbach's α's greater than 0.7, indicating good internal consistency. In addition, the mean and distribution of their inter-item correlation coefficients indicated homogeneity. Taken together, these results suggest that these subscales are likely to measure a single underlying construct (Field, 2009). In contrast, the Gregariousness towards Horses subscale had a Cronbach's α well below the threshold for acceptable internal consistency. This may be due to the fact that this subscale only comprises of three items, as Cronbach's α is negatively affected by the number of scale items (Cortina, 1993). However, the low mean and wide distribution of inter-item correlation coefficients also indicate potential multidimensionality in the scale (Simms and Watson, 2007). It therefore appears likely that there is heterogeneity in the underlying constructs measured by the scale. Indeed, for all 6 raters removing the item Q7: "Generally how dependable would you say this horse is?" resulted in an increase of Cronbach's α above the threshold for acceptable internal consistency. This might point to an issue with item selection for this scale. Therefore, the

Agreeableness, Neuroticism, Extraversion and Gregariousness towards People subscales show good internal consistency and homogeneity, reflecting the fact that all items on the subscales reflect the intended underlying personality construct. However, the Gregariousness towards Horses subscale may not be unidimensional and some items on that subscale may not accurately reflect this personality factor.

The inter-rater reliability analysis resulted in high ICC coefficients (ICC > 0.8) for the Agreeableness, Neuroticism, Extraversion and Gregariousness towards People subscales. This demonstrates good levels of agreement between raters compared to published thresholds for inter-rater reliability (Koo and Li, 2016). By comparison, intra-class coefficients ranging from 0.28 (4 raters, Agreeableness) to 0.53 (2 raters, Neuroticism) are reported for the human NEO Personality Inventory (McCrae and Costa, 1987) . While intra-class coefficients have not previously been used to assess the inter-rater reliability of equine personality questionnaires, preventing direct comparisons with the EPT, average intra-class coefficients of 0.62 and 0.79 have been reported for canine personality assessments (Gosling et al., 2003; Ley et al., 2009). Therefore, the first 4 subscales of the EPT show good inter-rater reliability compared to the published standards in human and domestic animal personality assessment. For three of those four subscales, and for most of the individual scale items making them up, primary caregivers achieved better inter-rater reliability than instructors. This was expected, as differential exposure to the target individual is known to affect inter-rater reliability: consistently being exposed to an animal in a particular context may limit the range of behaviours a judge has the opportunity to observe, and can therefore influence their perception of its personality (Funder et al., 1995; Gosling, 2001). Instructors were most familiar with the horses while they were being ridden, a relatively narrow context in which behavioural expression is reduced and largely placed under the control of the rider (Hall et al., 2008). However, differences in reliability between the two groups were relatively minimal. In addition, the ICC coefficients obtained by instructors remained well above published thresholds for acceptable agreement (Koo and Li, 2016), especially for the subscale level. Therefore, it appears that the restricted context in which they knew the horses, as well as the behavioural restrictions placed on ridden horses, did not significantly impede riding instructors' ability to reliably judge Agreeableness, Neuroticism and Extraversion. Overall, the Equine Personality Test therefore provides a highly reliable assessment of Agreeableness, Neuroticism, Extraversion, and Gregariousness towards People. Ratings are reliable even when provided with riding instructors who are familiar with the horses in a relatively narrow context. However, reliability is further improved when the ratings are provided by primary caregivers.

However, the ICC coefficient was low (ICC=0.498) for the Gregariousness towards Horses subscale, indicating poor inter-rater reliability (Koo and Li, 2016). This may be due to the fact that there was less variation in the scores for this factor, as ICC is sensitive to the amount of variation not only in the ratings but also in the sample (Hallgren, 2012). However, it seems more likely that this result reflects difficulty on the part of the raters to assess Gregariousness towards Horses traits reliably. Indeed, at the item level, items related to social behaviour towards other horses (Q3-5) also showed poor reliability. This might be due to the fact that the instructors, who made up the majority of the set of raters (k=3 out of 5), only knew the horses in a context where social behaviour is difficult to observe (Funder et al., 1995; Gosling, 2001). Indeed, in their informal email feedback, riding instructors self-reported difficulty in scoring items relating to behaviour towards other horses. This was not the case for primary caregivers, who observe the horses in a much wider set of circumstances, including when turned out in groups. However, rater familiarity was likely not the only factor driving the poor inter-rater reliability of the Gregariousness towards Horses subscale. While primary caregivers showed better agreement than instructors on this subscale, their ICC coefficients remained relatively low (ICC=0.562) and indicative of only moderate inter-rater reliability. This implies that even raters who had the opportunity to observe horses perform the relevant behaviours had difficulty in scoring those traits accurately. Funder (1995) suggests that some traits are inherently less observable, and therefore more difficult to rate than others. This could be the case here. It might also be that individual behavioural patterns on those traits show only limited stability across time and situations. Dominance rank has been shown to be linear in stable social groups, while affiliative relationships are developed with a network of preferred partners (Sigurjónsdóttir et al., 2003; Briard et al., 2015). Therefore, horses' tendencies to initiate aversive or affiliative social contacts might depend on the identity of the social partner present, making it difficult even for familiar raters to generalise their behaviour across situations. Gregariousness towards Horses was therefore assessed with only limited reliability by the panel of raters in this study, due to limited familiarity with the target horses but also to apparent difficulty in rating those traits. For studies concerned specifically with Gregariousness towards Horses, it may be preferable to use only primary caregivers as raters; however even in this case scores must be interpreted with caution.

ICC coefficients were generally high for both subscales and individual items in the test-retest reliability study, indicating that the scores given by raters using the EPT were consistent over time. Test-retest reliability was excellent for personality factors Neuroticism, Extraversion and Gregariousness towards People, and good for Agreeableness. While a few items on those scales had more limited test-retest reliability, the majority of items were rated consistently across time, suggesting that rater's perception of individual horse's level of expression of those traits remained constant across time. This is consistent with the idea that in adult animals, personality should reflect "temporally stable patterns of affect, cognition, and behaviour" (Gosling, 2008). However, while the ICC coefficient for the Gregariousness towards Horses subscale was also relatively high and showed acceptable consistency across time, all three items on the scale only had moderate test-retest reliability when taken individually. This suggests that, unlike the previous four, ratings on this subscale might show acceptable but limited temporal stability. As discussed above, this might be due to rater's difficulty to rate even familiar horses on those traits. Therefore, this chapter's results suggest that personality ratings on Agreeableness, Neuroticism, Extraversion and Gregariousness towards People collected using the EPT can be generalised beyond the time of collection. However, ratings on Gregariousness towards Horses only showed acceptable test-retest reliability and should be interpreted with caution when generalised over time. To the best of the author's knowledge, to date no other equine personality questionnaire has been assessed for test-retest reliability. This result therefore provides a benchmark for other questionnaires to be evaluated against.

It is interesting to note that the EPT seems to produce a particularly robust measure of Agreeableness. Given its links to behavioural compliance, this factor appears particularly relevant from an applied point of view in the equine industry (e.g. Graf, König von Borstel and Gauly, 2013). However, Agreeableness is generally the least internally consistent factor in human questionnaires evaluating the Five Factor Model (e.g. Scandell, 2000). Similarly, Momozawa et al. (2005) report their lowest internal consistency for their equine personality factor "Affability" (Cronbach $\alpha= 0.673$), which appears to be conceptually close to Agreeableness. In addition, Agreeableness is generally rated with limited reliability. In humans, this factor, along with Neuroticism, is identified as the least reliably rated of the 5 domains (John and Robins, 1993). While Neuroticism is assessed more reliably in animals, agreement on Agreeableness generally remains low (Gosling, 2001). By contrast, here comparable internal consistency and inter-rater reliability for Agreeableness, Neuroticism and Extraversion is reported. Agreeableness therefore seems to be measured in a comparatively robust way by the EPT. This could be due to high level of salience of these traits to horse handlers, given the use of the horse. Indeed, Gosling (2001) suggests that in animals, a given trait may be more easily observable in particular species or situations. As working or performance animals, horses are trained to perform locomotor responses to human cues, and are routinely exposed to and expected to tolerate aversive stimuli (Hall et al., 2008). This context might shine a particular light on Agreeableness traits while they may not be as easily observable in free ranging animals. In addition, handler safety depends largely on the handler's ability to predict a horse's level of compliance in those settings (Starling, McLean and McGreevy, 2016). As a result, Agreeableness traits may be particularly salient to handlers when forming their impression of a horse. Indeed, the most reliably rated EQP item ("Easy-going/Intolerant", ICC=0.875) belongs on the Agreeableness subscale. Other adjectives on the subscale relating to compliance with human cues ("Argumentative/Well-mannered", "Willing/Stubborn" and "Obedient/Wayward") all show high inter-rater reliability (ICC > 0.728). Agreeableness may therefore be more clearly and reliably measured in the horse than in other species. This could contribute making the horse a useful model species to study the biological bases of Agreeableness.

3.5 CONCLUSION

This study shows that in addition to predictive validity, four out of the five subscales of the EPT have satisfying internal consistency, inter-rater reliability and test-retest reliability. The questionnaire offers valid and reliable measures of the personality factors Agreeableness, Neuroticism, Extraversion, and Gregariousness towards People in equines. In particular, the EPT appears to produce a very robust measure of Agreeableness, a factor that is traditionally difficult to rate reliably in the human and animal personality literature but is highly relevant to equestrians. These subscales of the EPT therefore satisfy all four criteria of validity and reliability laid out in the psychometric literature. Thus, assessments of equine Agreeableness, Neuroticism and Extraversion carried out using the EPT can be used as a strong base from which to explore neurophysiological correlates of these personality factors. However, the Gregariousness towards Horses subscale proved to be problematic both in terms of internal consistency and reliability. Only primary caregivers showed acceptable if modest levels of agreement in their assessments of horses on this factor, and their assessment showed only limited consistency over time. Assessments on the Gregariousness towards Horses subscale should therefore be considered with caution.

Given the robustness of the first 4 factors of the EPT demonstrated here, in addition to its structural advantages, this questionnaire can be considered to result in the collection of valid and reliable personality data. It can therefore justifiably be used as a basis to investigate neurophysiological correlates of personality in the horse. In subsequent chapters, only personality scores provided by the three primary caregivers will be retained, owing to their improved reliability. They will be used to investigate the relationship between personality scores and (1) autonomic and HPA axis reactivity to challenges; (2) baseline HPA axis activity; and (3) striatal dopamine.

Chapter 4 Autonomic and HPA axis reactivity as a potential correlate of equine personality.

4.1 INTRODUCTION

Determining whether links exist between the reactivity of physiological stress systems and personality in the horse is of crucial applied importance. Personality assessment in the horse is generally seen as a means to safeguard human safety and equine welfare by selecting appropriate horses for demanding roles (König von Borstel, 2013). However, applied personality tests generally aim to select horses with limited behavioural reactivity to challenging situations (police horses: Pierard, McGreevy and Geers, 2017; sports horses: Lansade et al., 2016), with the unwritten assumption that their limited behavioural reactivity reflects a lack of stress. Accordingly, the majority of trait-based equine personality questionnaires are assessed for concurrent validity against behavioural outcomes (Anderson et al., 1999; Morris, Gale and Howe, 2002; Seaman, Davidson and Waran, 2002; Visser et al., 2003; Momozawa et al., 2003; Lloyd et al., 2007). This is also the case for the Equine Personality Test (EPT; Ijichi et al., 2013).

However, the literature on coping styles highlights that behavioural responses to challenging situations may not always accurately reflect the animal's internal states (Koolhaas et al., 1999, 2010). In the horse, emerging evidence suggests that overt behavioural responses to test situations or management procedures may not accurately reflect the intensity of the physiological stress response (Yarnell, Hall and Billett, 2013; Squibb et al., 2018; Munsters et al., 2013b). In particular, individuals who show little overt behavioural responses, and are therefore likely to score low for Neuroticism and Extraversion on the EPT (Ijichi et al., 2013), may nevertheless experience important physiological responses when exposed to a challenging situation (Yarnell, Hall and Billett, 2013; Munsters et al., 2013b). Therefore, questionnaires or test procedures that successfully identify individuals with limited behavioural reactivity (Pierard, McGreevy and Geers, 2017; Lansade et al., 2016) may not automatically result in the selection of individuals with equally low physiological stress sensitivity. Given the strong negative impact on welfare and performance of intense or prolonged physiological stress responses (Bartolomé and Cockram, 2016), it therefore seems crucial to ensure that personality assessment tools used to select horses for a role also have predictive validity for physiological responses to stressors.

The reactivity of the physiological systems involved in the adaptive stress response may be studied using non-invasive methods in the horse. Exposure to a stressor triggers an immediate, short-term response by the autonomic nervous system designed to sustain the fight, flight or freeze response (Bartolomé and Cockram, 2016). This response leads to physiological arousal and a shift in the balance of activity between the two branches of the autonomic nervous system, the sympathetic and parasympathetic (or vagal) nervous system (Von Borell et al., 2007). In order to prepare the body for acute, high intensity effort, the sympathovagal balance shifts towards sympathetic dominance (Von Borell et al., 2007). As both sympathetic and parasympathetic branches contribute to the regulation of cardiac function through the sinus node of the heart (Von Borell et al., 2007), this shift is reflected by cardiac parameters, such as heart rate and heart rate variability (Von Borell et al., 2007; Pierard et al., 2015; Stucke et al., 2015). During physiological arousal, the shift in sympathovagal balance towards sympathetic dominance results in an increase in heart rate and a decrease in heart rate variability (Von Borell et al., 2007). While they may be mediated by increased activity of the sympathetic nervous system, decreased activity of the parasympathetic nervous system, or a combination of both, increases in heart rate are thought to reflect in majority the influence of the sympathetic nervous system (Von Borell et al., 2007). By contrast, short term measures of heart rate variability such as RMSSD (the square root of the mean of the sum of the squares of differences between successive inter-beat intervals) are thought to reflect in majority the activity of the parasympathetic nervous system (Von Borell et al., 2007). In the horse, cardiac function is most frequently recorded using heart rate monitors (Stucke et al., 2015); this is generally due to practical and economic advantages, as they are affordable, portable, and fully non-invasive (Stucke et al., 2015). Heart rate and heart rate variability are well-established measures of physiological stress or arousal in the horse (Pierard et al., 2015). They are often used in equine studies to evaluate the physiological impact of husbandry procedures (e.g. Schmidt et al., 2010; Yarnell et al., 2013) or standardised test situations (e.g. Squibb et al., 2018; Safryghin, Hebesberger and Wascher, 2019).

Exposure to a stressor also triggers a response from the HPA axis, aiming to sustain the fight, flight or freeze response (Bartolomé and Cockram, 2016; Sapolsky, Romero and Munck, 2000). HPA axis activation triggers the release of cortisol, a circulating glucocorticoid which mediates responses aiming to maximise energy release towards the adaptive response (Bartolomé and Cockram, 2016). Accordingly, cortisol concentration may be used as a marker of HPA axis activity and physiological arousal in the horse (Pierard et al., 2015). Salivary cortisol offers a non-invasive measure of cortisol concentration in the horse (Peeters et al., 2011; Pierard et al., 2015). In the horse, exposure to stressful husbandry procedures such as road transport, clipping or training (Schmidt et al., 2010; Fazio et al., 2013; Yarnell, Hall and Billett, 2013) has been shown to lead to an increase in salivary cortisol. Therefore, the reactivity of the autonomic nervous system and HPA axis to stressors may be documented fully non-invasively in the horse, through the use of indicators of physiological arousal such as heart rate, RMSSD and salivary cortisol.

While routine husbandry procedures have been shown to induce autonomic and HPA axis responses in the horse (e.g. Schmidt, Möstl, et al., 2010; Fazio et al., 2013; Yarnell, Hall and Billett, 2013), these physiological systems may also be triggered using species-specific stress tests (Forkman et al., 2007). Compared to the use of naturally occurring stressors, reactivity tests can be standardised and therefore increase comparability within and across studies; they are also more easily quantifiable. Tests used to elicit physiological stress responses in a laboratory or controlled setting include the Trier Social Stress Test in humans (Kirschbaum, Pirke and Hellhammer, 1993), and the probe burying, raised maze or intruder paradigms in rodents (Koolhaas et al., 1999). By contrast, the most commonly used reactivity tests in horses are based on the horse's nature as a social, neophobic prey species (Wolff, Hausberger and Le Scolan, 1997). These include novel object tests, in which an unfamiliar object, either static or moving, is introduced in close proximity to the tested horse (Wolff, Hausberger and Le Scolan, 1997). Startle tests are also used, in which horses are exposed to a sudden visual and/or auditory stimulus at close range to trigger a startle response (usually an umbrella opening: Lansade, Bouissou and Erhard, 2008; Ijichi et al., 2013). Horses are generally tested alone and loose in the test arena, meaning that these tests generally do not include a handling component. However, handling tests may also be used, generally requiring the horse to be led over an unfamiliar surface (Wolff, Hausberger and Le Scolan, 1997; Ijichi et al., 2013; Visser et al., 2002) or past a novel object (König von Borstel et al., 2011), or to be exposed to a standardised version of a potentially stressful husbandry procedure (e.g. sham clipping: Yarnell, Hall and Billett, 2013).

Behaviour tests can be used to induce a mild fear response in the horse (Forkman et al., 2007). As such, they result in physiological arousal, including an autonomic and HPA axis response. Exposure to a novel object and startle tests have consistently been shown to induce an increase in heart rate (e.g. Visser et al., 2002; Christensen, Keeling and Nielsen, 2005; Mccall et al., 2006; Safryghin, Hebesberger and Wascher, 2019) and decrease in

RMSSD (Visser et al., 2002) in the horse. Likewise, handling tests such as a sham clipping procedure are associated with an increase in heart rate (Yarnell, Hall and Billett, 2013). The ability of standardised stress test to trigger HPA axis activation detectable through cortisol has not been demonstrated as consistently in the horse. Some studies do report a significant increase in plasma (Hada et al., 2001) or salivary (Yarnell, Hall and Billett, 2013) cortisol in response to exposure to a novel stimulus and a standardised sham clipping procedure, respectively. However other studies report null findings following exposure to comparable standardised tests (e.g. Noble et al., 2013; Villas-Boas et al., 2016; Ijichi et al., 2020). It should be noted that exposure to long-term stressors including musculoskeletal pain or inadequate management conditions impact both basal cortisol levels (Pawluski *et al.*, 2017) and HPA axis reactivity (Sauer *et al.*, 2019), thus potentially confounding the responses observed in standardised behavioural tests. Therefore, autonomic responses can reliably be induced in the horse using standardised fear tests such as a novel object test, a startle test, or a sham clipping procedure. HPA axis responses can also be induced using the same procedures, although a detectable activation may only be achieved with a higher intensity procedure and may be confounded by exposure to longer term stressors.

Sympatho-adrenal reactivity has been identified as a correlate of personality and coping style in human (Ormel et al., 2013) and rodent (Koolhaas et al., 2010) models, respectively (reviewed in Section 1.3). Preliminary evidence suggests these links may be present in the horse as well (e.g. Visser et al., 2003; Momozawa et al., 2003; Fazio et al., 2013; Bohák et al., 2017; reviewed in Section 1.4.4). However, a disconnect between behavioural and physiological reactivity to challenges may confound this relationship (Yarnell, Hall and Billett, 2013; Squibb et al., 2018; Munsters et al., 2013b). The predictive validity of equine personality assessment tools for physiological reactivity is an important outcome in terms of welfare and has not yet been established for the EPT. The aim of this study was therefore to investigate whether autonomic and HPA axis reactivity is a correlate of equine personality as measured by the EPT. To this end, horses were exposed to three test situations expected to induce short-term mild stress responses and commonly used in behaviour-based personality assessment: a novel object test, a startle test and a handling test (sham clipping). Mean heart rate, heart rate variability and salivary cortisol responses to the tests were recorded and compared with Equine Personality Test scores. Due to the links between Neuroticism and sensitivity to threat, it was hypothesised that a positive relationship would be observed between Neuroticism and SNS/HPA axis reactivity. In addition, due to the links proposed between Extraversion scores and coping style, it was hypothesised that Extraversion would be negatively linked with HPA axis reactivity. Agreeableness was also included in the analysis because it is likely that working horses are selected on this factor; however, it was hypothesised that this factor would not be linked with autonomic or HPA axis reactivity.

4.2 METHODS

4.2.1 HORSES AND MANAGEMENT

Data was collected from 23 horses (15 geldings, 8 mares; mean age 11.7±4.0 years) recruited from Brackenhurst Equestrian Centre. Horses remained in their normal management for the duration of the study. More details on horse demographics and management are available in Chapter 2.

4.2.2 REACTIVITY TESTS

4.2.2.1 Facilities

For the novel object and startle tests, baseline and recovery measurements were taken in the horse's home environment. These were a 3.5*4.1m stable in an American barn (n=20) or a paddock with field shelter shared with a companion horse (n=4). The test arena was a 7.3*9.2m crew yard within an American barn, surfaced with rubber matting and wood shavings (Figure 4.1a). It had three half board half railing walls and a full board wall with a window to the outside. The window was closed during the novel object test but open during the startle test. The test arena was familiar to all horses. In order to avoid social isolation during the behavioural tests, two companion horses were placed in stables across the aisle from the crew test arena. While within sight and hearing of the tested horse, they were at a distance deemed sufficient to minimize any risk of emotional contagion (Figure 4.1d).

The sham clipping test was carried out in the horse's home stables as described above (n=16) or in a familiar 2.5*3.7m cross-tie stall in an American barn (n=4), depending on the usual procedure followed to clip each horse. In both cases, companion horses were placed in adjacent stables within sight and hearing of the tested horse to avoid social isolation.

Figure 4.1 – The facilities used in the Novel Object and Startle tests: pictures of (a) the test arena, (b) the Novel Object, and (c) the umbrella used as the startling stimulus; (d) schematic of the layout of the test arena and adjacent stables, including those used to house companion horses for the duration of the tests.

4.2.2.2 Test procedures

Ahead of all tests, horses were caught and loosely tethered in their home stables with a headcollar and lead-rope. A saliva sample was taken to assess baseline salivary cortisol concentration; details and justifications of sampling methods are available below in Section 4.2.2.4.1. Horses were then fitted with a Polar RS800CX heart rate monitor (HRM) set to record continuous RR interval data (Christensen et al., 2014; Bohák et al., 2018). Similarly, further detail and justification on HRM use are available below in Section 4.2.2.3. Test specific procedures then started.

4.2.2.2.1 *Novel object test procedure*

Horses were led to the test arena in a headcollar and leadrope, then released into the test arena. They were then given 5 minutes to habituate to the arena with no interference from handlers (Lansade, Bouissou and Erhard, 2008); this phase is referred to thereafter as the "Habituation phase". 5 minutes were considered sufficient to allow the horses to acclimate to their new surroundings, as releasing horses into a familiar arena is not considered to be a stressor if there is no element of social isolation (Le Scolan, Hausberger and Wolff, 1997). After habituation was complete, the handler entered the test arena again. Horses was caught again and stood facing the arena door, away from the window. Test procedures then started.

The procedure for the novel object test was adapted from Lansade, Bouissou and Erhard (2008) and Ijichi et al. (2013). While the horse was facing away from the window, a novel object (yellow and green inflatable garden toy: Figure 4.1b) was introduced through it into the test arena. Once the window was closed again, the horse was released and the handler exited the test arena. The horse was then left free to interact with the novel object for 5 minutes, with no interference from experimenters (Lansade, Bouissou and Erhard, 2008; Ijichi et al., 2013). This phase is referred to thereafter as the "Test phase". After the end of the test phase, horses were caught, led back to their home stable and loosely tethered while recovery monitoring was carried out.

4.2.2.2.2 *Startle test procedure*

Horses were led to the test arena in a headcollar and leadrope, then released into the test arena. As in the Novel Object test, they were given 5 minutes with no interference from handlers to habituate to their surroundings, including a closed black automatic umbrella (Figure 4.1c) held by the experimenter through the open window (Lansade, Bouissou and Erhard, 2008). This phase is referred to thereafter as the "Habituation phase". After habituation was complete, a handler entered the test arena to catch the horse again and the test procedure started.

The procedure for the startle test was adapted from Lansade, Bouissou and Erhard (2008) and Ijichi et al. (2013). Horses were led to a start line 1m away from the window (Lansade, Bouissou and Erhard, 2008). Handlers could not remain in the test arena during the startle for health & safety reasons. A feed bucket containing a handful of concentrate pellets was therefore placed on the ground on the start line to ensure horses remained in a standardised position ahead of the startle. The handler then exited the test arena. While the horses were feeding from the bucket, the umbrella was opened, exposing the horses to a sudden visual and auditory stimulus. The distance from the window to the start line was such that the umbrella would open close to the horse's head but could not make contact. The umbrella was closed again after 10 seconds and removed from the window opening. Horses remained in the arena with no further interference for 5 minutes, with the opportunity to return to the feed bucket (Lansade, Bouissou and Erhard, 2008). This phase is referred to thereafter as the "Test phase". After the end of the 5 minutes test period, horses were caught, led back to their home stable and loosely tethered while recovery monitoring was carried out.

4.2.2.2.3 *Sham clipping test procedure*

The sham clipping procedure was adapted from Yarnell, Hall and Billett (2013). Horses were left loosely tethered and undisturbed for a minimum of 5 minutes in the familiar environment in which the test was carried out (home stable or cross ties) to record baseline state. This phase is referred to thereafter as "Baseline". After baseline measurements had been completed, a familiar handler entered the test area holding electric clippers with the blades removed. The sham clipping procedure (thereafter "Test phase") then began. The clippers were turned on and placed against 6 body sites for 1 minute each, in the following sequence: left shoulder, left flank, left hindquarter, right hindquarter, right flank, and right shoulder. The clippers were then turned off and the handler exited the test area.

4.2.2.2.4 *Recovery measures*

Following all tests, the heart rate monitor was turned off and removed. Saliva samples for salivary cortisol concentration analysis were taken 10, 15, 30 and 45 minutes after exposure to the stressor (Hall, Kay and Yarnell, 2014). These timings were chosen to reflect the 10 to 20-minute delay needed for plasma cortisol increases to be reflected in salivary cortisol (Peeters et al., 2011; Yarnell, Hall and Billett, 2013), as well as individual variations in the time needed to reach peak cortisol concentration following exposure to a stressor (Hall, Kay and Yarnell, 2014; Yarnell, Hall and Billett, 2013).

4.2.2.3 Autonomic responses to the test situations

RR intervals were recorded continuously using a Polar RS800CX heart rate monitor (Christensen et al., 2014; Bohák et al., 2018). This consisted of an elasticated electrode belt with a transmitter attached, and a receiver watch on which recordings were stored. The electrode belt with its transmitter was fastened around the thorax so that electrodes were on the left-hand side of the rib cage, with the lowest point of the electrode band aligned with the point of the elbow (Figure 4.2: Yarnell, Hall and

Billett, 2013; Ijichi et al., 2020). The horses' coat was soaked with warm water at the site of the electrodes (Squibb et al., 2018) and Spectra 360 electrode gel (Parker Laboratories) was applied to the electrode belt to improve conduction (Stucke et al., 2015). The watch receiver was fastened to the horse's headcollar so that it remained within close range of the transmitter to ensure continuous transmission (Figure 4.2: Squibb et al., 2018; Ijichi et al., 2020).

Figure 4.2 – Pictures showing the typical placement of the Polar RS800CX heart rate monitor on subjects. The elasticated sensor belt was placed around the thorax so that the lowest point of the electrode band aligned with the point of the elbow. The watch receiver was attached to the headcollar to ensure continuous transmission from the sensor.

Following completion of the tests, RR interval recordings were saved onto a PC through the Polar interface software. They were then processed using Kubios HRV Standard 3.1.0 (Department of Applied Physics, University of Kuopio, Finland). In accordance with the procedure used in previous equine studies, artefacts in the RR interval data were corrected using the Artifact Correction function with a custom threshold of 0.3 (Squibb et al., 2018; Ijichi et al., 2020; Ille et al., 2014).

Autonomic responses to the test situations were assessed using cardiac parameters (Von Borell et al., 2007). Mean heart rate (mean HR) was used as an indicator of sympathetic responses to the test situations (Von Borell et al., 2007). In addition, the root mean square of successive differences in RR intervals (RMSSD) was used as an indicator of rapid changes in cardiac activity, reflecting parasympathetic responses (Von Borell et al., 2007). Heart rate variability parameters vary with recording length (Task Force, 1996); mean HR and RMSSD were therefore analysed from a standardised recording duration for all horses, phases and tests to ensure comparability. Five minutes is the recommended duration to observe short term autonomic responses (Task Force, 1996) and this corresponded well with the timings of the habituation and test procedures used here. Therefore, for the Novel Object and Startle tests, mean HR and RMSSD were taken from the two 5-minute sections of RR recording corresponding to the Habituation phase and the Test phase (Table 4.1). For the Sham Clipping test, mean HR and RMSSD were taken from the two 5-minute sections of RR recording corresponding to the Baseline phase and the first 5 minutes of the Test phase (Table 4.1). For all three tests, the mean HR response to the test situation (ΔHR) was defined as the difference between mean HR during the test and mean HR during Habituation/at baseline (Table 4.1). Similarly, the RMSSD response to the test situation (ΔRMSSD) was defined as the difference between RMSSD during the test and RMSSD during Habituation/at baseline (Table 4.1).

4.2.2.4 HPA axis responses to the test situations

4.2.2.4.1 *Sampling and assay materials*

Salivettes (Sarstedt, UK) were adapted for use in horses by sewing a cotton thread through them which could be held by the handler to prevent swallowing (Yarnell, Hall and Billett, 2013). Saliva samples were taken by placing the Salivettes in the oral cavity for a minimum of 30 seconds (Yarnell, Hall and Billett, 2013). After sampling, Salivettes were kept on ice in a cool box for a maximum of 4 hours, then frozen at -20°C until the time of analysis (Yarnell, Hall and Billett, 2013; Pierard et al., 2015). Once defrosted, they were centrifuged at 3500RCF for 30 minutes to retrieve saliva. Centrifugation speed and duration was adapted to ensure an adequate sample was retrieved from all Salivettes. Salivary cortisol concentration was assayed using a commercial high-sensitivity enzyme immunoassay (EIA) kit (Expanded Range High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit, Salimetrics, USA). Although it is designed for use in humans, this assay kit has been validated for use in horses and is frequently used in equine studies (e.g. Shanahan, 2003; Ellis et al., 2014; Sauer et al., 2019; Mott, Hawthorne and McBride, 2020). A total of nine kits were used to process the samples. In accordance with the manufacturer's instructions, the kits were placed in a fridge at 2-8°C immediately upon reception and stored there until 2 hours before the assays were carried out (Salimetrics, 2019).

4.2.2.4.2 *General principles of EIA assays*

Cortisol present in the sample and cortisol conjugated with horseradish peroxidase (HRP) competes to bind onto a coating of anti-cortisol antibodies at the bottom of the microtitre wells. HRP is an enzyme that catalyses an oxidative reaction between hydrogen peroxide and 3,3',5,5'-Tetramethylbenzidine (TMB), a chromogenic substrate. After addition of a solution of TMB and hydrogen peroxide to the wells, this reaction produces a blue compound, TMB diimine. The reaction can be halted with the addition of methasulfonic acid, which turns TMB diimine yellow. The absorbance of the resulting solution can be read at 450nm using a spectrophotometer, and is inversely proportional to the concentration of cortisol in the sample being assayed (Volpe et al., 1998).

Six Standards (cortisol solutions of known concentrations supplied by the manufacturer) are used to convert absorbance values for each well into cortisol concentrations (μg/dL). In addition, three types of wells are used for quality control. High and Low Controls are cortisol solutions of unknown concentrations supplied by the manufacturers; for the assay to be valid their computed concentrations must fall within a range supplied by the assay documentation. Blank and Non-Specific binding wells are used to ensure the specificity of the assay. Blanks are wells coated with antibodies in which no sample containing cortisol is added, aiming to ensure that none of the assay reagents binds to the anti-cortisol antibodies. Non-specific Binding wells are wells that are not coated with antibodies, aiming to ensure that none of the assay reagents bind to wells themselves.

4.2.2.4.3 *EIA protocol*

A protocol supplied by the manufacturer was followed to carry out the EIA assay (Salimetrics, 2019). All procedures below are as described by Salimetrics (2019). In preparation for the assay, the microtitre plate and reagents were brought to room temperature for 1.5 hours. 500mL of assay buffer were prepared by diluting 5mL of Wash Buffer Concentrate tenfold in deionized water. A plate layout was drawn, including Standards, Controls, Blanks, Non-Specific Binding wells and samples; all were run in duplicate to account for assay variability and allow calculation of assay precision (Figure 4.3).

25µL of each Standard, Control and sample were pipetted into the appropriate wells, following the pre-determined plate layout (Figure 4.3). 25µL of pure assay diluent was added to the Zero and Non-Selective Binding wells. 15µL of enzyme conjugate (a solution containing cortisol conjugated with horseradish peroxidase) was then diluted into 24mL of assay diluent and 200µL of the diluted enzyme conjugate solution pipetted into each well. The plate was shaken for 5 minutes at 500rpm on an EIA plate rotator (PMS-1000i, Grant Instruments, UK), then left to incubate at room temperature for 1 hour. During the incubation period, the cortisol present in the samples and HRP-conjugated cortisol compete to bind onto the anti-cortisol antibodies coating the microtitre wells.

The contents of the plate were then emptied over a sink and the plate was washed 4 times with the diluted wash buffer, then blotted dry. 200µL of TMB Substrate solution, containing TMB and hydrogen peroxide, was pipetted into each well. Contact with HRP catalyses an oxidative reaction between those two substrates that produces a blue compound, TMB diimine.

The plate was mixed for 5 minutes at 500rpm on the plate rotator, then left to incubate for a further 25 minutes at room temperature in its foil pouch. 50µL of Stop Solution, containing methanesulfonic acid, were pipetted into each well. The plate was mixed for 3 minutes at 500rpm on the plate rotator, then read at an absorbance of 450nm in a plate reader (Multiskan FC, Thermo Scientific, US) within 10 minutes of adding the Stop Solution. For each plate, raw absorbance data and calculated concentration data were exported from the plate reader onto a PC for further analysis.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	Std 1	Std 1	Ctrl H	Ctrl H	S7	S7	S15	S15	S23	S23	S31	S31
B	Std 2	Std 2	Ctrl L	Ctrl L	S8	S8	S16	S16	S24	S24	S32	S32
C	Std 3	Std 3	S1	S1	S9	S9	S17	S17	S25	S25	S33	S33
D	Std 4	Std 4	S2	S2	S10	S10	S18	S18	S26	S26	S34	S34
E	Std 5	Std 5	S3	S3	S11	S11	S19	S19	S27	S27	S35	S35
F	Std 6	Std 6	S4	S4	S12	S12	S20	S20	S28	S28	S36	S36
G	Blank	Blank	S5	S5	S13	S13	S21	S21	S29	S29	S37	S37
H	NSB	NSB	S6	S6	S14	S14	S22	S22	S30	S30	S38	S38

Figure 4.3 – Layout of the 96 wells microtitre plates used for cortisol EIA assays, with (blue) the six Standards, (light green) Blanks, (dark green) Non-specific Binding wells, (light yellow) High Control, (dark yellow) Low Control and (grey) the 38 samples all assayed in duplicate.

4.2.2.4.4 Precision of assay

A total of 9 ELISA plates were used to assay the samples from the three behavioural tests, with 3 plates used for each test. Intra- and inter-assay coefficients of variation (CV) were calculated to evaluate the repeatability of the EIA assay results. The EIA kit supplier recommends considering intra-assay CV lower than 10% and inter-assay CVs lower than 15% as acceptable (Salimetrics, 2020). In the equine literature, a similar range of intra- and inter-assay CVs are considered acceptable when assaying salivary cortisol concentration (Ellis et al., 2014). For the Novel Object test, the intra-assay CV across all samples was 9.59% while the inter-assay CV between the three EIA plates was 6.31%. For the Startle test the intra-assay CV across all samples was 11.08%, while the inter-assay CV between the three EIA plates was 14.99%. For the Sham Clipping test, the intra-assay CV across all samples was 10.27%, while the inter-assay CV between the three EIA plates was 6.41%. Therefore, inter-assay CVs indicate good reliability, even though intra-assay CVs are close to the upper acceptable limit.

For each test, baseline and peak salivary cortisol concentrations were determined, defined as the concentration immediately before the test, and the maximum concentration between 10 and 45 minutes post-test respectively (Khoury et al., 2015) (Table 4.1). The salivary cortisol response to the test (ΔCORT) was then defined as the difference between peak salivary cortisol concentration post-test and salivary cortisol concentration at baseline (Khoury et al., 2015) (Table 4.1).

Table 4.1 – Summary table of all physiological measures taken to describe autonomic and HPA axis responses to the three reactivity tests.

Test situation	Physiological variable	Notation	Timing and context of measure
Novel Object test	Mean HR	$HR_{NO,Hab}$	Mean HR measured during the 5 minutes Habituation phase in the test arena, pre-test
		$HR_{NO,Test}$	Mean HR measured during the 5 minutes exposure to the Novel Object in the test arena
		ΔHR_{NO}	**Mean HR response to the Novel Object test: $\Delta HR_{NO} = HR_{NO,Test} - HR_{NO,Hab}$**
	RMSSD	$RMSSD_{NO,Hab}$	RMSSD measured during the 5 minutes Habituation phase in the test arena, pre-test
		$RMSSD_{NO,Test}$	RMSSD measured during the 5 minutes exposure to the Novel Object in the test arena
		$\Delta RMSSD_{NO}$	**RMSSD response to the Novel Object test: $\Delta RMSSD_{NO} = RMSSD_{NO,Test} - RMSSD_{NO,Hab}$**
	Salivary cortisol	$CORT_{NO,Bas}$	Salivary cortisol at rest in the home environment, before the start of all procedures
		$CORT_{NO,Peak}$	Maximum salivary cortisol from samples taken 10, 15, 30 and 45 minutes post-test
		$\Delta CORT_{NO}$	**Salivary cortisol response to the Novel Object test: $\Delta CORT_{NO} = CORT_{NO,Test} - CORT_{NO,Hab}$**
Startle test	Mean HR	$HR_{ST,Hab}$	Mean HR measured during the 5 minutes Habituation phase in the test arena, pre-test
		$HR_{ST,Test}$	Mean HR measured during the 5 minutes immediately post-startle in the test arena
		ΔHR_{ST}	**Mean HR response to the Startle test: $\Delta HR_{ST} = HR_{ST,Test} - HR_{ST,Hab}$**
	RMSSD	$RMSSD_{ST,Hab}$	RMSSD measured during the 5 minutes Habituation phase in the test arena, pre-test
		$RMSSD_{ST,Test}$	RMSSD measured during the 5 minutes immediately post-startle in the test arena
		$\Delta RMSSD_{ST}$	**RMSSD response to the Startle test: $\Delta RMSSD_{ST} = RMSSD_{ST,Test} - RMSSD_{ST,Hab}$**
	Salivary cortisol	$CORT_{ST,Bas}$	Salivary cortisol at rest in the home environment, before the start of all procedures
		$CORT_{ST,Peak}$	Maximum salivary cortisol from samples taken 10, 15, 30, and 45 minutes post-test

		$\Delta CORT_{ST}$	**Salivary cortisol response to the Startle test: $\Delta CORT_{ST} = CORT_{ST,Test} - CORT_{ST,Hab}$**
Sham Clipping	Mean HR	$HR_{CL,Bas}$	Mean HR at rest during the 5 minutes baseline, before sham clipping
		$HR_{CL,Test}$	Mean HR during the first 5 minutes of the sham clipping procedure
		ΔHR_{CL}	**Mean HR response to Sham Clipping: $\Delta HR_{CL} = HR_{CL,Test} - HR_{CL,Bas}$**
	RMSSD	$RMSSD_{CL,Bas}$	RMSSD at rest during the 5 minutes baseline, before sham clipping
		$RMSSD_{CL,Test}$	RMSSD during the first 5 minutes of the sham clipping procedure
		$\Delta RMSSD_{CL}$	**RMSSD response to Sham Clipping: $\Delta RMSSD_{CL} = RMSSD_{CL,Test} - RMSSD_{CL,Bas}$**
	Salivary cortisol	$CORT_{CL,Bas}$	Salivary cortisol at rest in the home environment, before the start of all procedures
		$CORT_{CL,Peak}$	Maximum salivary cortisol from samples taken 10, 15, 30 and 45 minutes post-test
		$\Delta CORT_{CL}$	**Salivary cortisol response to Sham Clipping: $\Delta CORT_{CL} = CORT_{CL,Test} - CORT_{CL,Hab}$**

4.2.3 **PERSONALITY DATA**

Personality scores were obtained for all horses in the sample using the Equine Personality Test (EPT: previously used and validated in Ijichi et al., 2013). This questionnaire uses subjective trait ratings by a familiar handler to assign a horse with a continuous score between 1 and 5 on five equine personality factors: Agreeableness, Neuroticism, Extraversion, Gregariousness towards People, and Gregariousness towards Horses. Here, ratings were provided for all horses by the same 3 primary caregivers, who had been familiar with the horses for a minimum of a year at the time of assessment. Questionnaires were scored for each rater individually, and for each horse the scores given by the 3 raters were then averaged to obtain the final set of personality scores used in the analysis. Further details on questionnaire items, scoring methods, and validity and reliability of the Equine Personality Test are available in Chapter 3 - Internal consistency and reliability of the Equine Personality Test.

4.2.4 **STATISTICAL ANALYSIS**

All statistical analysis was carried out using R (version 3.6.1; R Core Team, 2019). Figures were generated using the R package ***ggpubr*** (Alboukadel, 2020).

4.2.4.1 Relationship between physiological stress sensitivity and personality

Neither autonomic nor HPA responses were consistent across test situations (see Appendix A: Consistency of autonomic and HPA axis responses across test situations). Therefore, the responses to each test were considered separately, rather than as an average across the test situations. A correlation analysis was used to explore whether autonomic and HPA axis reactivity to challenges could be identified as a correlate of personality. The main physiological variables of interest for this chapter were the mean HR, RMSSD and salivary cortisol responses to each test (ΔHR, ΔRMSSD and ΔCORT), rather than the absolute value of those physiological parameters during the tests. This is because responses to species-specific challenges, rather than absolute values, are generally used in human and rodent studies documenting links between autonomic and HPA axis reactivity to challenges and personality in those species (reviewed in e.g. Chida and Hamer, 2008; Koolhaas et al., 2010; Ormel et al., 2013).

Personality scores, as well as the responses for each variable and test were tested for normality using Shapiro-Wilk tests (Field, 2009); most variables proved to vary significantly from the normal distribution. Spearman correlations were therefore used throughout to test the hypotheses linking physiological stress sensitivity to Neuroticism, Extraversion and Agreeableness. The Benjamini-Hochberg procedure was used to minimize the number of Type I errors associated with multiple testing, with the false discovery rate fixed at 10% (Benjamini and Hochberg, 1995). For each correlation, both uncorrected (p) and Benjamini-Hochberg adjusted (adj. p) p values are reported.

Correlation analyses are limited, as potential confounding factors cannot be accounted for in the analysis. For this reason, prior to settling for this approach, attempts were made to analyse the data using a modelling approach. For each physiological variable, a linear mixed model was to be used to explain values of the variables in the test as a function of test type, value of the variable at baseline and personality scores, with horse included as a random term. However, data exploration (Zuur, Ieno and Elphick, 2010) revealed that the data failed to meet some model assumptions and/or unsatisfactory distribution of the residuals indicating inadequate model fit were observed. This modelling approach was therefore abandoned, and simple correlations were used instead.

4.2.4.2 Physiological responses to the test situations

It has previously been hypothesised that relationships between physiological responses to stressors and personality traits such as fearfulness may only be apparent when the stressors induce a sufficiently strong physiological response (Christensen, Keeling and Nielsen, 2005). Therefore, the impact of the test situations on physiological variables was assessed, in order to inform the interpretation of the results of the correlation analysis between physiological responses and personality. This was done by comparing mean heart rate, RMSSD and salivary cortisol at baseline and after exposure to the stressors. For each pair of variables, Shapiro-Wilk tests were used to test the normality of the residuals, defined as the difference between the two paired variables (Field, 2009). Wilcoxon signed rank tests were used to compare baseline and post-test readings when residuals were not normal (Field, 2009). When they were, paired T Tests were used instead (Field, 2009).

4.2.5 ETHICAL APPROVAL

The protocol used for this study was reviewed by Nottingham Trent University's School of ARES Ethical Review Board. It received ethical approval on 21/05/2018 (project reference number ARE785). In accordance with the protocol, the yard manager confirmed the horses used were suitable to be exposed to the behavioural tests. An independent observer with investment in the study monitored the horses during both behavioural tests and had the authority to halt proceedings at any point should they feel it was necessary to safeguard welfare.

4.3 RESULTS

4.3.1 RELATIONSHIP BETWEEN PHYSIOLOGICAL STRESS SENSITIVITY AND PERSONALITY

4.3.1.1 Neuroticism

Neuroticism scores were not significantly correlated with either HPA axis or autonomic reactivity in any of the three test situations (Table 4.2).

4.3.1.1.1 *Mean heart rate*

The mean heart rate response to the test situation was not significantly correlated with Neuroticism in the Novel Object test (n=23, r= -0.071, adj. p= 0.747), the Startle test (n=23, r= 0.192, adj. p= 0.578) or the Sham Clipping test (n=20, r= -0.094, adj. p= 0.868).

4.3.1.1.2 *Heart rate variability*

The RMSSD response to the test situation was not significantly correlated with Neuroticism in the Novel Object test (n=23, r= -0.074, adj. p= 0.747), the Startle test (n=23, r= -0.190, adj. p= 0.578) or the Sham Clipping test (n=20, r= -0.103, adj. p= 0.868).

4.3.1.1.3 *Salivary cortisol concentration*

The salivary cortisol response to the test situation was not significantly correlated with Neuroticism in the Novel Object test (n=23, r= 0.076, adj. p= 0.747), the Startle test (n=23, r= -0.058, adj. p= 0.792) or the Sham Clipping test (n=20, r= -0.037, adj. p= 0.868).

Table 4.2 - Summary of Spearman rank correlations between Neuroticism scores and physiological changes in response to each of the three test situations. Both original and Benjamini-Hochberg adjusted (false discovery rate = 10%) p values are presented.

Test	Physiological response	N	Rho	p value	Adjusted p value
Novel Object	ΔHR	23	-0.071	0.747	0.747
	ΔRMSSD	23	-0.074	0.738	0.747
	ΔCORT	23	0.076	0.729	0.747
Startle test	ΔHR	23	0.192	0.381	0.578
	ΔRMSSD	23	-0.190	0.386	0.578
	ΔCORT	23	-0.058	0.792	0.792
Sham Clipping	ΔHR	20	-0.094	0.669	0.868
	ΔRMSSD	20	-0.103	0.641	0.868
	ΔCORT	20	-0.037	0.868	0.868

4.3.1.2 Extraversion

Extraversion scores were not significantly correlated with either HPA axis or autonomic reactivity in any of the three test situations (Table 4.3).

4.3.1.2.1 *Mean heart rate*

The mean heart rate response to the test situation was not significantly correlated with Extraversion in the Novel Object test (n=23, r= 0.061, adj. p= 0.782), the Startle test (n=23, r= -0.111, adj. p= 0.861) or the Sham Clipping test (n=20, r= -0.176, adj. p= 0.421).

4.3.1.2.2 *Heart rate variability*

The RMSSD response to the test situation was not significantly correlated with Extraversion in the Novel Object test (n=23, r= 0.084, adj. p= 0.782), the Startle test (n=23, r= 0.039, adj. p= 0.861) or the Sham Clipping test (n=20, r= -0.195, adj. p= 0.421).

4.3.1.2.3 *Salivary cortisol concentration*

The salivary cortisol response to the test situation was not significantly correlated with Extraversion in the Novel Object test (n=23, r= -0.153, adj. p= 0.782), the Startle test (n=23, r= 0.133, adj. p= 0.861) or the Sham Clipping test (n=20, r= -0.264, adj. p= 0.421).

Table 4.3 – Summary of Spearman rank correlations between Extraversion scores and physiological changes in response to each of the three test situations. Both original and Benjamini-Hochberg adjusted (false discovery rate = 10%) p values are presented.

Test	Physiological response	N	Rho	p value	Adjusted p value
Novel Object	ΔHR	23	0.061	0.782	0.782
	ΔRMSSD	23	0.084	0.703	0.782
	ΔCORT	23	-0.153	0.486	0.782
Startle test	ΔHR	23	-0.111	0.613	0.861
	ΔRMSSD	23	0.039	0.861	0.861
	ΔCORT	23	0.133	0.545	0.861
Sham Clipping	ΔHR	20	-0.176	0.421	0.421
	ΔRMSSD	20	-0.195	0.373	0.421
	ΔCORT	20	-0.264	0.224	0.421

4.3.1.3 Agreeableness

Agreeableness scores were not significantly correlated with either HPA axis or autonomic reactivity in any of the three test situations (Table 4.4).

4.3.1.3.1 *Mean heart rate*

The mean heart rate response to the test situation was not significantly correlated with Agreeableness in the Novel Object test (n=23, r= -0.272, adj. p= 0.314), the Startle test (n=23, r= -0.143, adj. p= 0.789) or the Sham Clipping test (n=20, r= 0.030, adj. p= 0.893).

4.3.1.3.2 *Heart rate variability*

The RMSSD response to the test situation was not significantly correlated with Agreeableness in the Novel Object test (n=23, r= 0.377, adj. p= 0.229), the Startle test (n=23, r= 0.121, adj. p= 0.789) or the Sham Clipping test (n=20, r= 0.316, adj. p= 0.425).

4.3.1.3.3 *Salivary cortisol concentration*

The salivary cortisol response to the test situation was not significantly correlated with Agreeableness in the Novel Object test (n=23, r= -0.202, adj. p= 0.356), the Startle test (n=23, r= 0.059, adj. p= 0.789) or the Sham Clipping test (n=20, r= -0.068, adj. p= 0.893).

Table 4.4 - Summary of Spearman rank correlations between Agreeableness scores and physiological changes in response to each of the three test situations. Both original and Benjamini-Hochberg adjusted (false discovery rate = 10%) p values are presented.

Test	Physiological response	N	r	p value	Adjusted p value
Novel Object	ΔHR	23	-0.272	0.210	0.314
	ΔRMSSD	23	0.377	0.076	0.229
	ΔCORT	23	-0.202	0.356	0.356
Startle	ΔHR	23	-0.143	0.516	0.789
	ΔRMSSD	23	0.121	0.582	0.789
	ΔCORT	23	0.059	0.789	0.789
Sham Clipping	ΔHR	20	0.030	0.893	0.893
	ΔRMSSD	20	0.316	0.142	0.425
	ΔCORT	20	-0.068	0.758	0.893

4.3.2 PHYSIOLOGICAL RESPONSES TO THE TEST SITUATIONS

4.3.2.1 Novel Object test

The Novel Object test did not have a significant effect on any of the physiological variables measured when compared to the Habituation period in the test arena (Figure 4.4).

The mean heart rate during the Novel Object test (mean ± SD: 50.46 ± 11.99 bpm) was not significantly different from the mean heart rate during habituation (mean ± SD: 47.82 ± 12.53 bpm) (Paired T test: t_{22}=0.97, p=0.342). In addition, RMSSD during the Novel Object test (mean ± SD: 72.23 ± 17.23 ms) was not significantly different from RMSSD during habituation (mean ± SD: 75.39 ± 26.75 ms) (Paired T test: t_{20}=-0.66, p=0.517).

Peak salivary cortisol concentration after exposure to the Novel Object test (mean ± SD: 0.144 ± 0.045 µg/dL) was not significantly different from salivary cortisol concentration at baseline immediately before the test (mean ± SD: 0.119 ± 0.032 µg/dL) (Wilcoxon test: W=159, p=0.135).

4.3.2.2 Sham clipping test

The Sham Clipping test did not have a significant effect on any of the physiological variables measured when compared to a baseline when loosely tethered in the stable (Figure 4.4).

The mean heart rate during the Sham Clipping test (mean ± SD: 36.97 ± 10.05 bpm) was not significantly different from mean heart rate at baseline immediately before the test (mean ± SD: 37.04 ± 7.83 bpm) (Wilcoxon test: W=97, p=0.722). In addition, RMSSD during the Sham Clipping test (mean ± SD: 87.23 ± 35.91 ms^2) was not significantly different from RMSSD at baseline immediately before the test (mean ± SD: 83.61 ± 40.18 ms^2) (Wilcoxon test: W=107, p=0.955).

Peak salivary cortisol concentration after exposure to the Sham Clipping test (mean ± SD: 0.099 ± 0.033 µg/dL) was not significantly different from salivary cortisol concentration at baseline before the test (mean ± SD: 0.088 ± 0.026 µg/dL) (Paired T test: t_{19}=-1.542, p=0.139).

4.3.2.3 Startle test

The Startle test did not have a significant effect on any of the physiological variables measured when compared to the Habituation period in the test arena (Figure 4.4).

The mean heart rate during the Startle test (mean ± SD: 53.01 ± 7.26 bpm) was significantly higher than the mean heart rate during habituation (mean ± SD: 47.82 ± 4.62 bpm) (Paired T test: t_{22}=2.87, p=0.009). In addition, RMSSD during the Startle test (mean ± SD: 65.19 ± 17.79 ms^2) was significantly lower than RMSSD during habituation (mean ± SD: 77.95 ± 22.82 ms^2) (Wilcoxon test: W=34, p=0.002).

Peak salivary cortisol concentration after exposure to the Novel Object test (mean ± SD: 0.078 ± 0.026 µg/dL) was not significantly different from salivary cortisol concentration at baseline before the test (mean ± SD: 0.067 ± 0.023 µg/dL) (Paired T test: t_{22}=1.51, p=0.144).

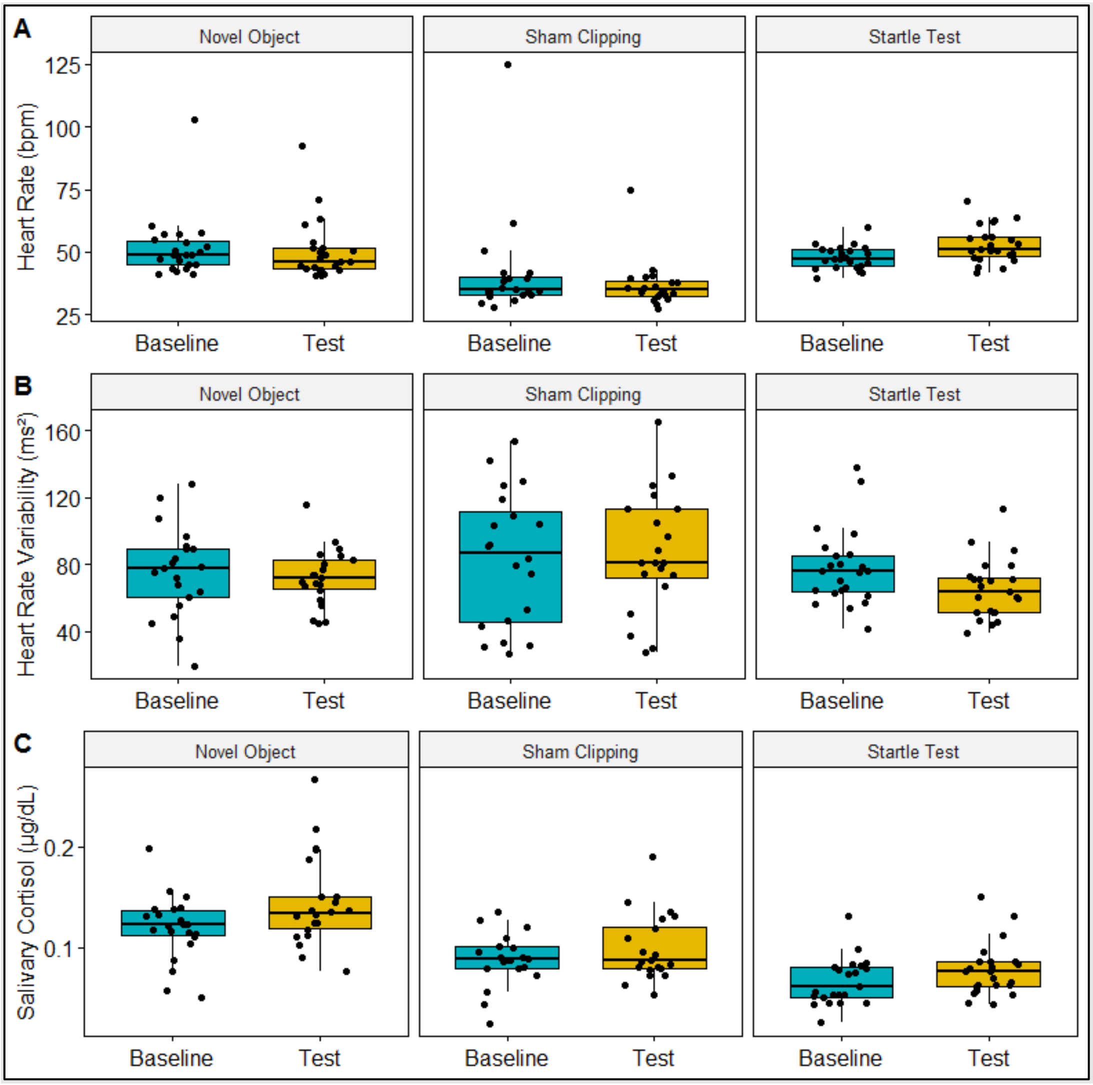

*Figure 4.4 - Boxplots comparing **(A)** Mean heart rate (bmp), **(B)** Heart rate variability (ms), and **(C)** Salivary cortisol concentration (µg/dL) at baseline and during the test, for each of the three test situations used.*

4.4 DISCUSSION

Autonomic and HPA axis reactivity to stressors has been associated with personality in both human and rodent models (Ormel et al., 2013; Koolhaas et al., 2010). However, in horses, only tentative evidence is currently available to support a link between personality dimensions such as Neuroticism and Extraversion, and the sensitivity of physiological systems to threat (Visser et al., 2003; Momozawa et al., 2003; Fazio et al., 2013; Bohák et al., 2017). This is an important outcome from an applied point of view as evidence suggests that the overt behavioural signs of stress predicted by those personality factors (Ijichi et al., 2013) may be a poor reflection of physiological states (Squibb et al., 2018). Therefore, clarifying the link between autonomic and HPA axis reactivity and personality is crucial to better understand whether horses selected for their desirable personality also have low physiological stress sensitivity, rather than a more passive behavioural expression of negative states. Therefore, this study aimed to explore whether the Neuroticism, Extraversion and Agreeableness as measured by the EPT were linked with physiological stress sensitivity. To this end, riding school horses (n=23) were exposed to mildly aversive behavioural test situations and their autonomic and HPA axis reactivity were monitored, through changes in heart rate variability and salivary cortisol respectively. The data did not support any of the hypothesised links between personality and autonomic and HPA axis reactivity. This suggests that personality dimensions that have been shown to reflect overt stress-related behaviour in the horse may nevertheless not give a clear insight into physiological stress reactivity.

The correlation analysis revealed no links between personality factors and autonomic or HPA axis reactivity in any of the test situations. The lack of relationship between Agreeableness and physiological stress reactivity is in line with the hypothesis poroposed for this factor. It is also coherent with previous results that evidenced no links between Agreeableness as measured by the EPT and behavioural responses to similar behavioural tests (Ijichi et al., 2013). Although the lack of relationship between physiological reactivity and Extraversion contradicts the hypothesis proposed for this factor, there are currently no comparable results in the equine literature that could be used to evaluate its reliability. The lack of relationship between physiological stress reactivity and Neuroticism contradicts theoretical frameworks of Neuroticism in other species (Ormel et al., 2013). However, in light of pre-existing evidence in the domestic horse, it is not wholly unexpected. While some studies do document a link between autonomic reactivity and Neuroticism-like equine personality traits (Visser et al., 2003; Momozawa et al., 2003; König von Borstel et al., 2011), it should be noted that none used the EPT to obtain their measure of trait-stress sensitivity. The discrepancies between the null finding presented here and those by Momozawa et al. (2003), Visser et al. (2003) and König von Borstel et al. (2011) could therefore be due to differences in the underlying construct measured by EPT-Neuroticism, as opposed to the anxiety/fearfulness constructs measured in those studies. Other studies suggest there may not be a consistent link between behavioural and physiological reactivity to stressors in the horse (Munsters et al., 2013b; Yarnell, Hall and Billett, 2013; Squibb et al., 2018). EPT-Neuroticism is assessed on the basis of behavioural reactivity to stressors (Ijichi *et al.*, 2013); therefore, the lack of relationship between physiological stress reactivity and Neuroticism documented here is coherent with these findings.

From an applied point of view, the lack of correlation between physiological responses to the test situations and personality scores suggests that the EPT cannot be used to predict the intensity of physiological responses to challenges, even though it has predictive validity for behavioural responses to similar challenges (Ijichi et al., 2013). If the EPT does not have predictive validity for physiological stress responses, then selecting horses for specific roles or lifestyles on the basis of the EPT may not be enough to reduce the risk of negative welfare consequences linked with prolonged or repeated activation of the physiological stress response (Khansari, Murgo and Faith, 1990; Sapolsky, Romero and Munck, 2000). Therefore, further research is needed to confirm the absence of a link between personality as assessed by the EPT and physiological stress reactivity, as it has important implications in terms of the practical applicability of the EPT.

It should be noted that this study presents a number of limitations. First, the test situations used elicited only a very mild aversive response from the horses in the sample. Salivary cortisol concentration was not significantly affected by any of the test situations, indicating that there was no detectable HPA axis response to the behavioural tests. This is in line with emerging findings suggesting that short-term exposure to mild stressors such as a novel or startling stimulus may not be sufficient to elicit measurable plasma cortisol responses (Noble et al., 2013; Villas-Boas et al., 2016; Minero, Zucca and Canali, 2006) or on salivary cortisol responses (Ijichi et al., 2020) in the domestic horse. Limited autonomic responses were also observed. As hypothesised and in line with previous findings in the horse (König von Borstel et al., 2011; Safryghin, Hebesberger and Wascher, 2019), the Startle test did prompt a significant increase in mean HR and decrease in RMSSD, consistent with physiological arousal (Von Borell et al., 2007). However, in contradiction with hypotheses, the Novel Object test and Sham Clipping procedure failed to induce changes in autonomic activity. This is unexpectedas significant autonomic responses indicative of an arousal/stress response have been reported when horses are exposed to novel stimuli (e.g. Visser et al., 2002; Christensen, Keeling and Nielsen, 2005; Mccall et al., 2006; Safryghin, Hebesberger and Wascher, 2019), or a sham clipping procedure (Yarnell, Hall and Billett, 2013). Therefore, the test situations triggered an unexpectedly low physiological stress response in the sample recruited for this project. It has been suggested that relationships between heart rate reactivity and behavioural responses indicative of fearfulness may only be apparent when the test situations induce a sufficiently strong heart rate response (Christensen, Keeling and Nielsen, 2005). Therefore, the limited aversiveness of the test situations may contribute to explain the unexpected lack of relationship observed between physiological responses and Neuroticism.

The limited physiological responses to the test situations observed may be linked in part with ethical considerations that impacted the experimental design and sample used in this study. For ethical reasons, only very mild stressors could be used, with control measures in place to limit the impact of the procedures. In particular, companion horses were placed in neighbouring stables to limit any impact of social isolation in all test situations. However, the presence of a calm individual has been shown to reduce fear responses to a novel object in a naïve horse (Christensen et al., 2008). Although companions were not placed in the test arena alongside the tested horse, nearby social support might have helped buffer responses to the test situation regardless. Interestingly, recent evidence suggests that social buffering decreases the fear response to a novel, but not sudden stimulus (Ricci-Bonot et al., 2021). This might contribute to explain why autonomic responses were observed in the Startle test, but not the Novel Object test. In addition, the selection of horses used for these studies did not include any deemed likely to be distressed by the situation. In particular, horses known to be non-compliant with clipping were excluded, unlike in the study conducted by Yarnell et al. (2013) which included both compliant and non-compliant horses. When these groups were considered separately, compliant horses showed a mild decrease, rather than an increase in mean HR during the procedure (Yarnell, Hall and Billett, 2013), in line with the findings of this chapter. Standardised tests were chosen here over opportunistic data collection as they allow for better standardisation of measurements. However, this result suggests that the impact of ethical restrictions on the aversiveness of the test situations may limit their relevance in the context of this study. Therefore, future research may instead consider the use of opportunistic data collection during routine husbandry procedures known to trigger strong physiological responses (e.g. trailer loading: Shanahan, 2003; travelling: Fazio et al., 2013) in order to ethically document physiological reactivity to more highly aversive situations.

Sample characteristics may also have contributed to explain the limited physiological responses observed. The sample recruited for this project consisted of mature riding school horses belonging to a university teaching and research herd. As such, these horses were routinely exposed to a wide range of novel stimuli as part of their normal working lives. They also regularly took part in novelty testing for research purposes. By contrast, the samples used in most studies documenting much clearer physiological responses to similar test situations (e.g. Visser et al., 2002; Christensen, Keeling and Nielsen, 2005) consisted of young horses up to 2 years of age, that had received little or no previous handling. This difference in sample demographic and life experience may contribute to explain the difference in responses observed. Indeed, older horses in some samples have been shown to display less pronounced behavioural responses when exposed to a novel stimulus, although their physiological responses were not tested (Bulens et al., 2015). In addition, horses may be capable of a degree of generalisation after habituation to a novel object, which dampens behavioural and physiological responses to further novelty (Christensen, Zharkikh and Chovaux, 2011). While the horses were naïve to the specific object used in this study, it cannot be ruled out that some, or all, had previously been exposed to an object similar enough to enable generalisation. Therefore, sample characteristics might in part contribute to explain the unexpected lack of physiological response to the Sham Clipping and Novel Object test, and the comparatively mild response to the Startle test.

In addition to the limiting factors linked with ethical considerations and sample demographics, this study is limited by the reliability of the methods used to assess physiological responses to the test situations. Firstly, electrocardiogram (ECG), rather than

HRM, is considered the gold standard measure for inter-beat intervals (Von Borell et al., 2007; Parker et al., 2009). However, due to their affordability HRMs are most frequently used to monitor cardiac responses in equine welfare studies (Stucke et al., 2015). Similarly, affordability and availability guided the choice of using HRMs in the present study. Nevertheless, only limited evidence is available to support the validity of HRM data against the gold standard of ECG in equines (Ille et al., 2014; Lenoir et al., 2017; Parker et al., 2009). Accordingly, the RR traces obtained in the present study appeared very artefact-rich, although the absence of an ECG trace made it impossible to identify artefactual measures with certainty (Von Borell et al., 2007). Good agreement has been found between HRM and ECG data when the horse is stationary (Parker et al., 2009; Ille et al., 2014). However, agreement is more limited when the horses is in movement, and decreases as movement increases (Parker et al., 2009; Lenoir et al., 2017). Here, the horses were restrained during the Sham Clipping test but were free to move around a small test arena (7.3*9.2m) in the Novel Object and Startle test. Therefore, although movement was limited to a walk due to the confined space (Stucke et al., 2015), it may have contributed to the high density of data points likely to be artefacts.

In addition to movement, a number of other factors outside of the author's control may have contributed to the limited quality of the HRM traces. Parker et al. (2009) recommends that the coat should be freshly clipped before use of an HRM in order to maximise conductivity between the skin and the electrodes. However, this could not be done here for ethical reasons. In addition, recent evidence in humans suggests that a higher percentage of body fat negatively influences the reliability of a HRM when compared to ECG data (Hernández-Vicente et al., 2021). Although no body condition score data was formally collected, the body condition of the horses in the sample was generally above optimal, which may have affected the quality of the HRM traces obtained. Although the data was corrected using the correction protocol most often described in equine studies ("Artefact correction" in Kubios software, set to “Custom: 0.3”, used in e.g. Schmidt et al., 2010c; Ijichi et al., 2020; Squibb et al., 2018; Ille et al., 2014), visual inspection suggested some errors may have been retained in the corrected traces used for HR/HRV analysis. This was especially the case for type 4/5 errors, where artefactual measurements concerned a section of the recording rather than a single beat (Parker et al., 2009). Errors retained in the corrected trace may limit the degree to which the values of HRV parameters reflect actual autonomic activity in the sample, and therefore may have confounded potential relationships between HRV parameters and personality. As portable ECG devices become more readily available, this study should be replicated using this more reliable measure of inter-beat interval.

Recent evidence also calls into question the reliability of the salivary cortisol analysis performed in this study. The Salimetrics ELISA assay kit used in this chapter has frequently been used to assess salivary cortisol concentrations in equine studies (e.g. Shanahan, 2003; Ellis et al., 2014; Sauer et al., 2019; Mott, Hawthorne and McBride, 2020). However, a recent publication has questioned its reliability for salivary cortisol concentrations below 18ng/mL in the horse (=1.8 µg/dL) (Sauer et al., 2020), in accordance with findings in human research (Bae et al., 2016). Although the salivary cortisol concentrations found in the present study were within the assay range as described by the manufacturer (0.012-3.000 µg/dL: Salimetrics, 2019), they were below this newly identified threshold by 2 to 3 orders of magnitude. In particular, Sauer et al. (2020) suggest that for concentrations below the 18ng/mL threshold, there is no correlation between the concentration obtained via ELISA assay and that obtained through the use of liquid chromatography-tandem-mass spectrometry (LC-MS/MS), which is considered the gold standard in human salivary research (Bae et al., 2016; Miller et al., 2013). This suggests that the rank order of concentrations observed here may not reliably reflect the rank order of actual HPA axis reactivity in the sample, thus making it difficult to accurately identify a correlation between personality and HPA axis reactivity. The results of this chapter should therefore be treated with caution. Due to the low cortisol concentrations observed both at baseline and in response to the tests, any studies aiming to replicate these findings should consider assaying salivary cortisol concentration by LC-MS/MS instead.

4.5 CONCLUSION

This chapter aimed to establish whether autonomic and HPA axis reactivity could be identified as a physiological correlate of personality in the horse, in accordance with human and rodent models. Mean HR, heart rate variability and salivary cortisol were used to evaluate autonomic and HPA axis reactivity in three standardised test situations commonly used to trigger physiological arousal: a novel object test, a startle test and a handling test (sham clipping procedure). Contrary to hypotheses, a correlation analysis failed to reveal any link between physiological stress reactivity and the conserved personality dimensions Agreeableness, Neuroticism and Extraversion. From an applied point of view, this suggests that the EPT may not have suitable predictive validity for physiological stress responses to challenges. However, the present study is limited by the low aversiveness of the test procedures used and the reliability of the methods used to monitor physiological stress reactivity. Therefore, further research is needed in order to confirm the findings of the present study. Crucially, future studies should be conducted on a more diverse and representative sample in which previous experience could be quantified and included as a factor in the analysis. In addition, the use of test situations eliciting a stronger response should be considered. Given the strong physiological responses documented during routine management practices (e.g. loading or travelling), opportunistic data collection, rather than the use of behavioural tests for the purpose of research, may enable researchers to ethically document responses to more aversive situations. While this chapter focussed on the reactivity to stressors of the autonomic nervous system and HPA axis, the baseline activity of those systems, and in particular of the HPA axis, has also been linked with aspects of personality. The following chapter will therefore focus on establishing whether the chronic activity of the HPA axis, rather than its reactivity, can be identified as a correlate of personality in the horse.

Chapter 5 Chronic HPA axis activity as a potential correlate of equine personality

5.1 INTRODUCTION

Chapter 4 suggests that point-in-time measures of cortisol reactivity to stressors may not be suitable to investigate links between hypothalamic–pituitary–adrenal (HPA) axis function and equine personality. By contrast, in humans, basal cortisol levels are positively associated with fearfulness and the tendency to experience psychological stress (Zilioli et al., 2015; Montoya et al., 2012). Basal cortisol levels are also associated with the balance between reward and punishment sensitivity, with lower basal cortisol levels associated with reduced punishment sensitivity and increased reward dependency (Honk et al., 2003). Due to this, basal cortisol levels in humans are also associated with traits such as social dominance, social aggression and empathy (Montoya et al., 2012; Honk et al., 2003; Zilioli et al., 2015; Mehta et al., 2015). Similar patterns reminiscent have also been described in animals in the context of the coping style framework. Indeed, reactive copers, who are characterised by passive responses to stressors and low rates of social aggression (Koolhaas et al., 1999), have elevated baseline cortisol concentration compared to active copers (Koolhaas et al., 2010). Therefore, basal or chronic cortisol levels may prove more relevant than point-in-time measures when exploring links between HPA axis function and equine personality.

Measuring chronic HPA axis activity has been problematic until recent years. Cortisol levels are commonly assayed from a number of matrices including blood, saliva, urine and faeces (Mormède et al., 2007). However, these only offer short- to mid-term information on cortisol levels. Plasma and saliva offer point-in-time measures of HPA axis activity, reflecting circulating cortisol concentration at or immediately before the time of sampling (Mormède et al., 2007; Peeters et al., 2011b; Yarnell, Hall and Billett, 2013). While these measures accurately capture responses to acute stressors, they are subject to fluctuations due to circadian rhythms (Mormède et al., 2007; Bohák et al., 2013), environmental disturbances (Schmidt et al., 2010b; Becker-Birck et al., 2013; Peeters et al., 2013) and sampling stress (Mormède et al., 2007). Urine and faeces, on the other hand, offer a cumulative overview of cortisol excretion over up to 48 hours for faeces in the horse (Palme, 2012). However, both remain sensitive to fluctuations linked with exposure to short-term stressors (Merl et al., 2000; Berghold, Möstl and Aurich, 2007). While repeated sampling using these matrices can provide a cumulative estimate of HPA axis activity over a longer period, this measure does not represent a true index of chronic HPA axis activity (Meyer and Novak, 2012) and has disadvantages such as repeated exposure to sampling stress and high analytical costs. Therefore, blood, saliva, urine and faeces provide limited value in measuring chronic HPA activity.

In the last two decades, numerous studies have shown that cortisol can also be extracted and assayed from hair (Raul et al., 2004; Gow et al., 2010; Davenport et al., 2006). Cortisol is incorporated passively from the bloodstream into the hair matrix during hair growth and remains stable in the hair shaft, thus providing an overview of chronic HPA axis activity during the period of hair growth (Henderson, 1993; Gow et al., 2010). Hair cortisol offers an overview of chronic HPA axis activity over the period of hair growth from a single measure, with no impact of short-term fluctuations due to circadian rhythms or acute stressors (Meyer and Novak, 2012; Russell et al., 2012). In addition, hair sampling is fully non-invasive, and the kinetics of cortisol incorporation into the hair matrix ensure that sampling stress cannot confound the cortisol concentration assayed (Russell et al., 2012). Cortisol concentration remains stable in hair samples stored for months or years at room temperature, allowing for long-term storage before analysis (Wennig, 2000; Davenport et al., 2006). Due to these numerous advantages, hair cortisol is increasingly used as an index of chronic HPA axis activity, with a large body of literature now associated with this biomarker in both human and non-human animals (e.g. reviewed in Meyer and Novak, 2012; Russell et al., 2012; Heimbürge et al., 2019).

The most common use of hair cortisol is as a biomarker of chronic stress (Russell et al., 2012). Hair cortisol has been used in a wide array of wild, captive, and domestic animal species to assess the impact of stressors as varied as anthropogenic disturbances (Ewacha et al., 2017; Jacobson et al., 2017; Martin and Réale, 2008), social conflict (Yamanashi et al., 2013, 2018), or husbandry practices (Stradaioli et al., 2017; Schubach et al., 2017). However, in the absence of differences in exposure to environmental stressors, hair cortisol has also successfully been used to explore the relationship between basal cortisol levels and personality traits. Hair cortisol concentration is negatively associated with novelty-seeking phenotype in Vervet monkeys (Laudenslager et al., 2011). It is also positively related with behavioural reactivity to standardised tests such as an auditory startle in dogs (Siniscalchi et al., 2013) or a Human Intruder Test in Rhesus macaques (Hamel et al., 2017). Hair cortisol is also linked with the tendency to initiate aggression in chimpanzees, although the direction of the relationship differs by sex (Yamanashi et al., 2016), and positively associated with Sociability in the common Marmoset (Inoue-Murayama et al., 2018). Finally, hair cortisol is positively associated with a Docility index in Eastern Chipmunks (Martin and Réale, 2008). Therefore, in continuity with findings in humans linking basal cortisol levels with the balance of sensitivity to punishement and reward (Honk et al., 2003), hair cortisol has been linked with a range of personality traits associated with behavioural reactivity to stressors and the tendency for compliant or agonistic social behaviour.

To date, no links between equine personality and HPA axis function estimated through point-in-time measures of cortisol levels have been established. Indeed, baseline cortisol levels as measured through a single plasma sample taken at were found to be unrelated to personality (Anderson et al., 1999). In addition, HPA axis reactivity to stressors, as measured through the salivary cortisol response, was not correlated with personality (Chapter 4). By contrast, recent evidence suggests that hair cortisol concentration is significantly negatively correlated with personality factors Dominance, Anxiousness, and Excitability (Sauveroche et al., 2020). At the trait level, the later two factors appear conceptually close to EPT-Neuroticism and EPT-Extraversion, respectively (reviewed in Section 1.4.3). These factors also appear consistent with the traits linked with baseline HPA axis activity in other species, including behavioural reactivity to stressors (Anxiousness: Ormel et al., 2013), coping style (Excitability: Koolhaas et al., 2010) and the tendency for agonistic social behaviour (Dominance: Montoya et al., 2012). Therefore, these emerging results suggest that hair cortisol may be a useful tool to probe relationships between HPA axis activity and equine personality (Sauveroche et al., 2020).

In addition to allostatic load and underlying differences in neurophysiology, other factors including age, sex, sampling location and hair pigmentation have been shown to impact hair cortisol concentration in mammals (reviewed in e.g. Heimbürge et al., 2019). Age has a consistent pattern of impact on hair cortisol across species, with elevated hair cortisol concentrations at birth that decrease in early life, but no relationship between age and hair cortisol concentration in adult samples (e.g. cows: González-de-la-Vara et al., 2011; pigs: Heimbürge et al., 2020; pig-tailed macaques: Grant et al., 2017; baboons: Fourie et al., 2015; polar bears: Neuman-Lee et al., 2017). Emerging findings in young foals (Comin et al., 2012;

Montillo et al., 2014) and adult samples (Sauveroche et al., 2020; Gardela et al., 2020) appear to confirm this pattern in the horse. Sampling location has also been shown to consistently impact hair cortisol concentration in a wide range of species (e.g. pigs and cattle: Heimbürge et al., 2020; brown bear: Macbeth et al., 2010; chimpanzees: Carlitz et al., 2015; baboons: Fourie et al., 2016; kangaroos: Sotohira et al., 2017). In horses, studies report differences in hair cortisol concentrations between permanent and coat hair samples (Sauveroche et al., 2020), between permanent hair samples from the mane and the tail (Duran et al., 2017), and among coat hair samples from different anatomical locations (Banse et al., 2020). Therefore, emerging evidence in the horse suggests that age may not confound hair cortisol concentration in adult samples, but that sampling location should be taken into account in analyses.

The impact of sex and coat colour on hair cortisol concentration is not as well established (Heimbürge, Kanitz and Otten, 2019). In some species no impact of sex is reported (e.g. Asiatic black bear: Malcolm et al., 2013; brown bear: Macbeth et al., 2010; dog: Bennett and Hayssen, 2010; Roth et al., 2016; Packer et al., 2019; orang-utan: Carlitz et al., 2014; pigs and cattle: Heimbürge et al., 2020). However, in other species elevated hair cortisol is reported in males (e.g. American black bear: Lafferty et al., 2015; chimpanzee: Yamanashi et al., 2016; Jacobson et al., 2017; coyote: Schell et al., 2017) or in females (e.g. polar bear: Bechshøft et al., 2011; Neuman-Lee et al., 2017; baboon: Fourie et al., 2016; Rocky mountain goat: Dulude-de Broin et al., 2019). In horses, most studies report no impact of sex on hair cortisol (Comin et al., 2012; Montillo et al., 2014; Duran et al., 2017; Sauveroche et al., 2020), although one reports lower hair cortisol concentrations in mares (Prinsloo et al., 2019). However, most equine studies reporting on mixed-sex samples use young animals that have not yet reached sexual maturity (Comin et al., 2012; Montillo et al., 2014; Duran et al., 2017). Finally, very little data is available in the horse on the impact of hair colour on hair cortisol concentration. In other species, lower hair cortisol concentrations are generally reported in black compared to non-black hair (e.g. cattle: González-de-la-Vara et al., 2011; Burnett et al., 2014; chimpanzees: Yamanashi et al., 2013; dogs: Bennett and Hayssen, 2010), although the opposite has also been reported (Heimbürge et al., 2020). Gardela et al. (2020) report no differences in hair cortisol concentration between the coat hair of grey and bay horses. However, it should be noted that this study comprises a very small sample size, and that the pigmentations compared are different from those used in other species. Therefore, further research is needed to consolidate current knowledge of the impact of sex and coat colour on hair cortisol concentration in the horse.

Chronic HPA axis activity has been linked with personality traits relating to stress-sensitivity and levels of social aggression in human and non-human animals. However, the links between basal cortisol levels and personality remain unclear in the horse. Hair cortisol is emerging as a reliable measure of chronic HPA axis activity, and has been linked with personality in a number of non-human species. Therefore, the aim of this study was to investigate the relationship between subjectively rated personality and basal cortisol concentration in the horse, measured through hair cortisol concentration. Sex, age, sampling location and hair colour were included in the analysis to account for their potential confounding impact on hair cortisol. Based on results in the horse (Sauveroche et al., 2020) and other species (Honk et al., 2003; Koolhaas et al., 2010), it was hypothesised that hair cortisol concentration would be negatively related with Neuroticism and Extraversion, but would not be significantly related with the other personality factors. In line with emerging findings in the adult horse (Duran et al., 2017; Banse et al., 2020; Sauveroche et al., 2020), it was hypothesised that sex, age and coat colour would not have a significant impact on hair cortisol, while sampling location would significantly impact hair cortisol concentration.

5.2 METHODS

5.2.1 HORSES

Hair samples were collected from 24 horses housed at the Brackenhurst Equestrian Centre (8 mares, 16 geldings; mean age: 13.4 ± 3.8 years). At the time of hair sample collection (September 2019) horses were coming towards the end of their summer conditioning program. During this period the horses are brought back into work by the Brackenhurst Equestrian Centre staff after the summer break, using a standardised exercise program. All horses were kept under similar management with minor tailoring to individual requirements (e.g. feed supplements provided where relevant). They were kept in groups on grass pasture at night and during weekends and were brought in to their home stables on week days. The horses had not been used for teaching in the 3 months preceding data collection. None had been exposed to significant stressors or experienced disease or medication in the months preceding hair sampling. Further details about samples demographics, management and exercise regime during the summer conditioning program are available in Chapter 2.

5.2.2 PERSONALITY ASSESSMENT

Personality was assessed using the Equine Personality Test (EPT: Ijichi et al., 2013). This questionnaire uses subjective trait ratings by a familiar handler to assign a horse with a continuous score between 1 and 5 on five equine personality factors: Agreeableness, Neuroticism, Extraversion, Gregariousness towards People, and Gregariousness towards Horses. Here, ratings were provided for all horses by the same 3 primary caregivers. All raters had been familiar with the horses for a minimum of a year at the time of assessment. Questionnaires were scored for each rater individually, and for each horse the scores given by the 3 raters were averaged to obtain the final personality score used in the analysis. Further details on questionnaire items, scoring methods, and validity and reliability of the Equine Personality Test are available in Chapter 3.

5.2.3 HAIR CORTISOL CONCENTRATIONS

5.2.3.1 Hair samples collection

Most hair cortisol studies in animals, including horses, sample from coat hair using a shaving-reshaving technique (Meyer and Novak, 2012). However, here mane hair was deemed more appropriate. This is because mane hair is not seasonally shed and could therefore be used to reflect HPA axis activity over a longer period than coat hair. In addition, rate of growth may be more accurately determined in permanent hair, allowing for a more precise estimate of the time period being assessed. Finally, equestrians are generally reluctant to carry out procedures that alter the visual characteristics of their horses. The aesthetic impact of a shaving-reshaving collection technique would likely prove to be an obstacle to the uptake of the method in an industry setting. In order to ensure the industry relevance of the technique developed here, permanent mane hair sampling was therefore preferred. While cutting manes is not well regarded, mane pulling is commonly undertaken under traditional management. Mane hairs were therefore pulled rather than shaved and hair samples including hair roots obtained.

Horses were caught in their home stables and loosely tethered using a headcollar and leadrope. Manes were pulled following the procedure routinely used as part of horse care at Brackenhurst Equestrian Centre. The handler held on to a strand of mane hair, combed the remaining hair up towards the neck using a mane comb, then wrapped the strand around the comb and pulled sharply downwards. Samples were collected from three locations along the mane: near the withers, halfway up the neck and near the poll. Once collected, samples were placed in sealed paper envelopes labelled with the horse's name and sample location and stored at room temperature until processing and analysis. One horse had a hogged mane. For this subject, hair samples could only be collected near the withers and pole but not at the midpoint of the neck.

5.2.3.2 Sample preparation

Previously published studies investigating hair cortisol concentrations in equines based their cortisol extraction and assay procedures on protocols developed by Koren et al. (2002) and Macbeth et al. (2010) for hair cortisol analysis in rock hyraxes and grizzly bears, respectively (Comin et al., 2012; Duran et al., 2017). In contrast, the laboratory procedures described thereafter were based on a published protocol for hair cortisol extraction and assay in human and monkey hair by Meyer et al. (2014). This protocol was chosen for its level of detail and because it is widely referenced in the wider hair cortisol literature. Minor amendments were made to the published protocol to account for differences in laboratory equipment available; they are highlighted and justified below. Several experimenters were involved in the laboratory analysis. The majority of procedures were carried out by the author (AJ), with assistance from JB and KG (see Table 5.2 for details of the work undertaken by each experimenter).

5.2.3.2.1 *Sample fragmentation*

Hair roots were removed to avoid potential contamination by circulating cortisol (Duran et al., 2017). Only the top 3cm of hair from the root were conserved for analysis because cortisol concentration has been shown to decrease along the length of the hair shaft in equine mane hair (Duran et al., 2017). Where hairs of different colours were present within a sample, they were separated into distinct samples, with only the largest one processed. Once fragmented, hair samples were placed into labelled 15mL screw-cap polypropylene centrifuge tubes.

5.2.3.2.2 *Sample washing and drying*

Sweat or sebum from the hair surface are potential sources of external cortisol contamination and must be removed prior to hair cortisol extraction (Davenport et al., 2006). Methanol and isopropanol are the two most commonly used wash solvents in hair cortisol analyses (Davenport et al., 2006; Kroshko et al., 2017). Methanol removes surface contaminants more efficiently (Pötsch and Moeller, 1996); however it is also more likely to penetrate the hair shaft and affect the hair cortisol concentration detected (Eser et al., 1997). Therefore, samples were gently washed using high performance liquid chromatography grade isopropanol.

5mL of isopropanol were pipetted into each tube. Samples were then rotated at 30rpm for 3 minutes before the isopropanol was decanted into a waste container, taking care not to lose any of the sample. Samples were washed 3 times instead of the 2 recommended in Meyer et al. (2014) as the hair was heavily coated with sebum and dirt, and visible contamination was still present after the second wash (Macbeth et al., 2010). This was deemed appropriate as washes only have limited impact on the retention of cortisol in the hair matrix (Davenport et al., 2006; Macbeth et al., 2010). The samples were then left to dry in a fume cabinet for 3 days, until the isopropanol was fully evaporated.

5.2.3.2.3 *Sample grinding*

Samples were finely ground using a bead mill (Bead Mill 24, Fisher Scientific, USA). A bead mill was used in order to maximise the chances of the samples yielding enough cortisol to be detected by the EIA assay. More finely ground samples yield higher hair cortisol concentrations due to increased surface area (Yamanashi et al., 2016b); in particular, processing the hair with a bead mill rather than surgical scissors yields higher cortisol concentrations (Burnett et al., 2014). Using a precision balance and tweezers, 60mg of hair from each sample were transferred into labelled reinforced 2mL tubes with screw caps (Fisher Scientific, US). The weight of each sample was recorded in grams down to 4 decimal points. Three 2.4mm metal beads were added to each tube in order to grind the samples in the bead mill. Tubes were then loaded onto the finger plate of the bead mill. In accordance with health and safety procedures, they were secured by screwing on the tube holder, engaging the 'locked' position and closing the glass window. After processing in the bead mill, the glass window was opened, the tube holder unlocked and unscrewed, and the tubes unloaded. The reinforced tubes containing the powdered hair were then set aside ready for cortisol extraction; the metal beads used for grinding remained in the tubes.

Grinding time and speed had to be adjusted up from Meyer et al. (2014) to account for differences in bead mill performance. Samples from the poll were processed first and ground in 4 successive bursts at increasing speeds with results checked in between each burst (Table 5.1). This showed that samples needed to be processed for 10 minutes at 6m/s to obtain satisfactory powdering. The rest of the samples were subsequently processed in a single burst for 9 minutes and 59 seconds at 6m/s. Because the end result was similar, it was assumed that the differences in grinding protocols between samples would not affect the amount of cortisol eventually detected. The grinding process pushed some of the powdered hair into the tube caps. This methodological difficulty, which could not be solved, made it difficult not to lose any of the sample when opening the tube, therefore potentially affecting the samples weight.

Table 5.1 – Grinding speeds and times tested on the 24 samples taken from the poll in order to obtain satisfactory grinding, accounting for differences in bead mill performance from the reference protocol (Meyer et al., 2014).

Grinding time	Grinding speed	Result
2 minutes	3m/s	No visible effect on hair samples.
+ 2 minutes	6m/s	Some hair powdered; majority of the sample unaffected.
+ 2 minutes	6m/s	Approximately half of the sample powdered; some hair strands remain completely unaffected.
+ 6 minutes	6m/s	Most of the sample powdered; a few longer strands remain.

5.2.3.3 Cortisol extraction

5.2.3.3.1 *Methanol extraction*

The solvent used to extract cortisol from the hair matrix was methanol, a lower alcohol in which steroids such as cortisol are highly soluble (Pötsch and Moeller, 1996). Methanol penetrates deeply within the hair matrix, ensuring maximum cortisol yield (Eser et al., 1997; Davenport et al., 2006). Samples were processed in 3 batches on 3 consecutive days, using identical methodology (Table 5.2). For health and safety reasons, all procedures that could result in exposure to methanol fumes were carried out under a fume hood. Under the fume hood, 1.5mL of high-performance liquid chromatography grade methanol was pipetted into each tube. Samples were then placed on a rotator to incubate at room temperature for 22 hours with constant rotation at 30rpm. The tubes were then centrifuged at 10 000rpm for

10 minutes in a Flowgen bioscience mini centrifuge. Under the fume hood, 1.0mL of supernatant was then pipetted into labelled 2mL Eppendorf tubes, taking care not to disturb the pelleted hair at the bottom of the tubes.

5.2.3.3.2 *Solvent evaporation*

The methanol solvent was then evaporated under the fume hood, using a block heater and sample concentrator (Stuart SBH130D/3 and SBHCONC/1, Cole-Palmer, USA) to accelerate the evaporation process. The Eppendorf tubes were uncapped and placed on the block heater set at 40°C. The sample concentrator was used to blow a flow of nitrogen gas over the open Eppendorf tubes in order to help displace the methanol vapour and accelerate evaporation. It was set so that the needles directing the nitrogen flow were just through the necks of the Eppendorf tubes but did not dip into the methanol and cortisol solution. Samples were left in the device for approximately an hour, until the tubes were fully dry. Regular visual checks were performed to evaluate the progress of evaporation.

5.2.3.3.3 *Sample reconstitution*

Samples were immediately reconstituted by pipetting 0.2mL of EIA assay diluent into each Eppendorf tube (Meyer et al*.,* 2014). The lowest dilution recommended in this protocol was used because hair cortisol concentration is generally low and cortisol concentrations in the reconstituted samples were unlikely to exceed the EIA kit's upper sensitivity limit. After addition of the assay buffer, samples were homogenized for approximately 5 seconds using a vortex mixer (Lab Dancer, IKA), then frozen at -20°C awaiting EIA assay.

5.2.3.4 Enzyme immunoassay

A commercial high-sensitivity enzyme immunoassay (EIA) kit (Expanded Range High Sensitivity Salivary Cortisol Enzyme Immunoassay Kit, Salimetrics, USA) was used to assay the cortisol concentration in the reconstituted samples. This kit was the same as used in Chapter 4 and has originally been developed to assay salivary cortisol concentration. However, it has been validated for hair cortisol assays (Davenport et al., 2006) and is one of the more frequently used commercial EIA kits in hair cortisol studies (Albar et al., 2013; Kroshko et al., 2017). Two kits were used to process the samples (Table 5.2). In accordance with the manufacturer's instructions, the kits were placed in a fridge at 2-8°C immediately upon reception and stored there until 2 hours before the assays were carried out. The protocol followed to assay cortisol concentrations in the reconstituted samples was the one supplied by the manufacturer (Salimetrics, 2019). Full details of this protocol, as well as a summary of the general principals of EIAs, are available in Sections 4.2.2.4.2 and 4.2.2.4.3.

Intra- and inter-assay coefficients of variation (CV) were calculated to evaluate the reliability of the EIA assay results. The EIA kit supplier recommends considering intra-assay CV lower than 10% and inter-assay CVs lower than 15% as acceptable (Salimetrics, 2020). In the equine literature, a similar range of intra- and inter-assay CVs are considered acceptable when assaying salivary cortisol concentration (Ellis et al., 2014). Here, the intra-assay CV across all reconstituted samples was 8.54%. The inter-assay CV between the two EIA plates was 6.4%. Results can therefore be considered as reliable.

5.2.3.5 Conversion of cortisol concentration in reconstituted samples to hair cortisol concentration

Concentrations in the reconstituted samples (µg/dL) and weights of hair samples (mg) were used to calculate hair cortisol concentrations in the original hair samples (pg/mg) using the following formula:

$$Hair\ Cortisol\ Concentration = 3000 * \frac{Concentration\ in\ reconstituted\ sample}{Weight\ of\ hair\ sample}$$

This formula was adapted from Meyer et al. (2014) by plugging in to their equation the volumes of methanol added to the powdered hair, supernatant recovered after methanol extraction and assay diluent added to reconstitute the samples used in the present study.

Table 5.2 – Experimental procedures carried out as part of the hair cortisol laboratory analysis. The table shows instances where samples were processed in separate batches and details the role of all experimenters in carrying out the analysis. Separate batches are including on separate lines. Samples from different horses were numbered consecutively from 1 to 26 and sample location is identified by the letters T (pole), B (withers) and M (midpoint of the neck).

Procedure	**Batches**	**Experimenter**
Hair cutting	1M to 24M, 1T to 3T, 1B to 3B	AJ
	5T to 24T, 5B to 24B	JB
	25T, 25M, 25B, 26T, 26M, 26B	KG
Sample washing	1T to 22T	AJ
	23T to 26T, 1M to 19M	AJ, KG
	20M to 26M, 1B to 15B	AJ, KG
	16B to 26B	AJ, KG
Sample weighing		AJ
Sample grinding	1T to 26T	AJ
	1M to 26M	AJ
	1B to 26B	AJ
Methanol addition	1T to 26T, 1M to 17M	AJ
	18 M to 26 M, 1B to 22B, 25B, 26B	KG
	23B, 24B	AJ
Supernatant recovery	1T to 26T, 1M to 17M	AJ
	18 M to 26 M, 1B to 22B, 25B, 26B	AJ
	23B, 24B	AJ
Solvent evaporation	1T to 26T, 1M to 17M	AJ
	18 M to 26 M, 1B to 22B, 25B, 26B	AJ
	23B, 24B	AJ
Sample reconstitution	1T to 26T, 1M to 17M	AJ, KG
	18 M to 26 M, 1B to 22B, 25B, 26B	AJ, KG
	23B, 24B	AJ
EIA assay	Plate 1: 1B to 26B, 1M to 15M	AJ
	Plate 2: 17M to 26M, 1B to 26B	AJ

5.2.4 STATISTICAL ANALYSIS

All statistical analysis was carried out in R (version 3.6.1: R Core Team, 2019). Briefly, a generalized linear mixed model (GLMM) was used to model hair cortisol concentration as a function of the 5 personality factors and 4 potential confounding factors. Model selection was used to determine which variables to retain in order to obtain the best fitting model. Figures were generated using the R package ***ggplot2*** from the ***tidyverse*** collection of packages (Wickham et al., 2019)

5.2.4.1 Data exploration

Prior to fitting a model, the data was explored (Zuur, Ieno and Elphick, 2010). There were no zeros or missing values in the response or explanatory variables. One statistically significant outlier was removed from the response variable (HCC=15.28pg/mg; Grubb's test: $G = 6.261$, $U = 0.432$, $p < 0.000$); there were no outliers in the explanatory variables. The response variable was continuous, strictly positive, homogenous, but not normally distributed (Shapiro-Wilk test: $W = 0.902$, $p < 0.000$). There was no collinearity (all $r < 0.6$) or multicollinearity (all $VIF < 3.5$) between explanatory variables. Interactions between explanatory variables were explored graphically; no clear evidence of interaction was found and therefore no interaction terms were included in the model.

Group sizes were balanced for all categorical variables except hair colour (Black: n=36; Chestnut: n=7; White: n=27). In order to balance the group sizes for this variable, hair colours were regrouped as 'Black' and 'Non-black'. This grouping is coherent with previous findings in dogs, in which hair cortisol concentration was shown to differ in black compared to non-black coat colours (Bennett and Hayssen, 2010).

5.2.4.2 Model formulation

The response variable was continuous, strictly positive, homogenous and non-normal. There was dependency in the model due to the repeated measures taken from the same individual at the different sample locations. Therefore, a Gamma generalized linear mixed model (GLMM) with log link was used to model hair cortisol concentration as a function of the ovariates. The fitted model (referred to thereafter as "the full model") took the form:

$$HCC_{ij} \sim Gamma\,(\mu_{ij}, \phi)$$

$$E\left(HCC_{ij}\right) = \mu_{ij} \text{ and } var\left(HCC_{ij}\right) = \frac{\mu_{ij}^2}{\phi}$$

$$log\left(\mu_{ij}\right) = \eta_{ij}$$

$$\eta_{ij} = \beta_1 + \beta_2 \times agree_{ij} + \beta_3 \times neur_{ij} + \beta_4 \times extr_{ij} + \beta_5 \times gregP_{ij} + \beta_6 \times gregH_{ij} + \beta_7 \times age_{ij} + \beta_9 \times sex_{ij} + \beta_{10} \times location_{ij} + \beta_{11} \times colour_{ij} + horse_j$$

$$horse_j \sim N(0, \sigma_{horse}^2)$$

Where HCC_{ij} is hair cortisol concentration in sample i collected from *horse* j, assuming a gamma distribution with mean μ and precision ϕ. The variables $agree_{ij}$, $neur_{ij}$, $extr_{ij}$, $gregP_{ij}$, $gregH_{ij}$ and age_{ij} were continuous covariates representing Agreeableness, Neuroticism, Extraversion, Gregariousness towards People, Gregariousness towards Horses and age of the horse, respectively. The variables sex_{ij}, $location_{ij}$ and $colour_{ij}$ were categorical covariates and represented sex of the horse, sampling location and hair colour of the sample, respectively. $Horse_j$ was included as a random intercept in the model to account for the repeated measures taken at different sampling locations within each individual. The model was fitted using the *glmer()* function from the ***lme4*** package (Bates et al., 2019).

5.2.4.3 Model selection

Manual backward selection was used in order to identify the best fitting plausible model. Manual iterations of the base R function *drop1()* (R Core Team, 2019) were used to identify the covariates that could be removed from the model in order to improve the fit. For a given model specification, *drop1()* fits all possible models with one single covariate removed to the data and returns the Akaike information criterion (AIC). The AIC estimates the quality of fit of each model: a lower AIC indicates a better fitting model. For each iteration of *drop1()*, the covariate whose deletion resulted in the lowest AIC (the best fitting model) was removed. Iterations were repeated until removing further variables resulted in a higher AIC, i.e. a worse fitting model. The model obtained through manual backward selection is referred to thereafter as "the final model".

5.2.4.4 Model validation

The fit of the final model was assessed through graphical means by plotting the residuals against both the fitted values and the retained covariates. There were no patterns within the residuals and the fit of the model was therefore considered satisfactory.

5.2.5 ETHICAL APPROVAL

This study received ethical approval from the ARES Ethical Review Board (project reference ARE795). Care was taken to protect welfare during data collection. There was no interruption of normal management for the horses in the sample. The procedure used for pulling manes is part of standard management at Brackenhurst Equestrian Centre. In addition, personality data was collected fully non-invasively and involved no direct interaction with the horses.

5.3 RESULTS

5.3.1 MODEL SELECTION

Iterations of *drop1()* resulted in the successive removal of covariates age_{ij}, $colour_{ij}$, $extr_{ij}$, sex_{ij}, $neur_{ij}$, $gregP_{ij}$, and $gregH_{ij}$. However, covariates $location_{ij}$ and $agree_{ij}$ were retained in the best fitting model. The final model therefore took the form:

$$HCC_{ij} \sim Gamma\,(\mu_{ij}, \phi)$$

$$E\left(HCC_{ij}\right) = \mu_{ij} \text{ and } var\left(HCC_{ij}\right) = \frac{\mu_{ij}^2}{\phi}$$

$$log\left(\mu_{ij}\right) = \eta_{ij}$$

$$\eta_{ij} = \beta_1 + \beta_2 \times agree_{ij} + \beta_3 \times location_{ij} + horse_j$$

$$horse_j \sim N(0, \sigma_{horse}^2)$$

This model had an AIC of 187.99, improved down from 195.00 for the initial full model.

5.3.2 IMPACT OF COVARIATES ON HAIR CORTISOL CONCENTRATION

Both covariates retained in the final model had a significant impact on hair cortisol concentration (Figure 5.1). There was a significant positive association between Agreeableness score and hair cortisol concentration (Table 5.3: p=0.01). In addition, there was a weak but significant effect of sampling location on hair cortisol concentration. Hair cortisol concentration was significantly higher at the poll than at the mid-point of the neck (Table 5.3: p=0.03). It also showed a non-significant tendency to be higher at the withers than at the mid-point of the neck (Table 5.3: p=0.07).

Table 5.3 - Summary of a Gamma GLMM to model hair cortisol concentration as a function of sampling location and Agreeableness. Samples from different horses were fitted as random intercepts. $N_{obs} = 71$.

	Estimate	**Std. Error**	**t**	**p**
(Intercept)	0.602	0.255	2.360	0.0183*
Agreeableness	0.172	0.067	2.569	0.0102*
Location$_{(Poll)}$	0.123	0.058	2.112	0.0347*
Location$_{(Withers)}$	0.106	0.059	1.805	0.0711

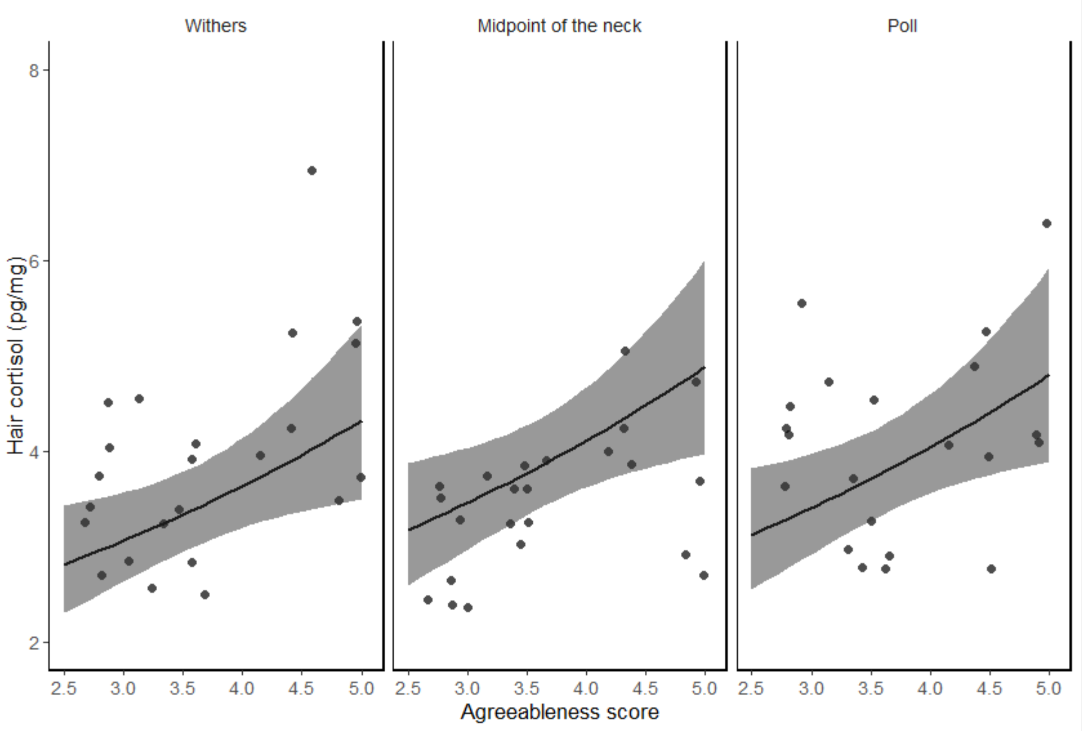

Figure 5.1 - Fitted values for hair cortisol concentration against Agreeableness scores for the three sampling locations along the mane, modelled using a Gamma GLMM. Grey bands indicate 95% confidence intervals around the fitted line. Black circles are observed values for hair cortisol concentration.

5.4 DISCUSSION

Basal cortisol levels have been linked with aspects of personality such as fearfulness and social aggression in a number of human and non-human species. However, in horses the link between subjectively rated personality and chronic HPA axis activity has not yet been well documented. To this end, personality and chronic HPA axis activity were assessed in 24 riding school horses kept under similar management, through subjective trait rating and hair cortisol concentration respectively. A GLMM was used to model hair cortisol concentration as a function of personality factors and four potential confounding factors identified in the horse and other species. A single confounding factor was found to affect hair cortisol; in addition, a significant negative association between Agreeableness and hair cortisol concentration was found.

5.4.1 IMPACT OF CONFOUNDING FACTORS

Age, sex and hair colour were not retained as explanatory variables of hair cortisol concentration during model selection. The absence of relationship between age and hair cortisol was expected and is consistent with previous results reported in adult individuals, in the horse (Gardela et al., 2020; Sauveroche et al., 2020) and in other species (Heimbürge et al., 2019). Similarly, the absence of impact of sex on hair cortisol concentration is as hypothesised and consolidates the findings of Sauveroche et al. (2020) in a second mixed-sex samples of adult horses. While sex differences in hair cortisol are reported in other species (e.g. American black bear: Lafferty et al., 2015; chimpanzee: Yamanashi et al., 2016; Jacobson et al., 2017; coyote: Schell et al., 2017; polar bear: Bechshøft et al., 2011; Neuman-Lee et al., 2017; baboon: Fourie et al., 2016; Rocky mountain goat: Dulude-de Broin et al., 2019), this suggests these differences may not be found in the horse and that sex may not need to be controlled for when assessing correlates of hair cortisol. It should be noted, however, that the male horses in the mixed-sex sample used in this study were all geldings. Therefore, this result may not be applicable in a sample that includes stallions. Overall, these results suggest that age and sex do not need to be controlled for when assessing correlates of hair cortisol in samples of adult mares and geldings.

The absence of impact of hair colour on hair cortisol concentration is coherent with the only result available in the horse to date (Gardela et al., 2020), although it contradicts findings in other species (bovine: González-de-la-Vara et al., 2011; Burnett et al., 2014; chimpanzee: Yamanashi et al., 2013; dog: Bennett and Hayssen, 2010). It should be noted that this finding suffers from limitations: group sizes for the three colours were relatively small, and the regrouping performed prior to applying the model in order to balance group sizes may not have been biologically relevant for the species. In addition, the analysis was carried out across individuals, while most studies reporting an impact of hair colour on hair cortisol focus on within-individual differences (González-de-la-Vara et al., 2011; Burnett et al., 2014; Yamanashi et al., 2013). Further research is therefore needed to confirm that hair colour does not significantly impact hair cortisol in the horse, as this result is inconsistent with findings in other species.

In accordance with hypotheses, sampling location was retained in the model as a covariate with a significant impact on hair cortisol concentration. Compared to levels at the midpoint of the neck, hair cortisol was significantly higher at the poll and tended to be higher at the withers. While Banse et al. (2019) have shown that hair cortisol concentration differs between close anatomical locations, to the best of the author's knowledge this is the first evidence that hair cortisol may vary between locations along the mane. The reasons behind those variations are unclear. They may be due to differential levels of vascularisation in surrounding tissues at the different sampling sites. In addition, exposure to mechanical irritation and repeated washing have also been shown to affect hair cortisol concentration

(Hamel et al., 2011; Salaberger et al., 2016), and mane hair at the different sampling sites may be differentially exposed to those. Unfortunately, because samples from the poll and the withers and those from the midpoint of the neck were prepared by different experimenters (see Table 5.2), this result could also be due to differences in sample preparation techniques or precision. Despite this limitation, this result underlines the importance of rigorously standardising hair sampling location even within an anatomical structure in future protocols, to ensure that hair cortisol concentrations are comparable across individuals.

5.4.2 IMPACT OF PERSONALITY FACTORS

Contrary to hypotheses, Neuroticism was not retained as an explanatory factor of hair cortisol concentration in the final model. The absence of link between Neuroticism and basal cortisol levels suggests that trait stress-sensitivity may not be reflected in chronic HPA axis activity in the horse. This finding is unexpected and may contradict previous results in the horse that link hair cortisol with the equine personality factor Anxiousness (Sauveroche et al., 2020). However, it should be noted that the personality assessment tool used by Sauveroche et al. (2020) is different from the one used in the present study, which could explain the discrepancy in findings. Despite similarities at the trait level, Anxiousness (Lloyd et al., 2007) and EPT-Neuroticism (Ijichi et al., 2013) have not formally been established to measure the same underlying construct. In addition, the use of a simple correlation analysis in Sauveroche et al. (2020) may not have enabled researchers to account for the impact of potential confounding factors such as breed and management practices, that may affect both personality and hair cortisol concentration (Lloyd et al., 2008; Sauveroche et al., 2020). Together with the results from Chapter 4, this finding suggests that Neuroticism scores were not associated with HPA axis function in the sample used in this project. This finding differs from those documented in humans (reviewed in Ormel *et al.*, 2013), and further research is needed to clarify the link between Neuroticism and the baseline activity and reactivity of the HPA axis in the horse.

Extraversion was not retained as an explanatory factor of hair cortisol in final model. This contradicts the hypothesis formulated here on the basis of emerging findings in the horse (Sauveroche et al., 2020). As discussed above, the discrepancies between the findings reported in this chapter and in Sauveroche et al. (2020) may be due methodological differences. Nonetheless, the results of this chapter appear more closely aligned with reports in other species. For instance, to date no links have been reported between Extraversion and hair cortisol levels in humans (Steptoe, Easterlin and Kirschbaum, 2017; Rietschel et al., 2017). In addition, two studies in bovines failed to identify a link between hair cortisol and excitability, a facet of Extraversion (Lockwood et al., 2017; Cooke et al., 2017). Therefore, more research is needed to clarify those emerging results on a potential link between Extraversion and basal cortisol levels in the horse.

Gregariousness towards People and Gregariousness towards Horses were not retained as explanatory variables of hair cortisol concentration during model selection. The absence of link between basal cortisol levels and the two Gregariousness factors is consistent with the hypotheses formulated. Indeed, while basal cortisol levels have been linked with the tendency to express social aggression in humans and animals (Montoya et al., 2012; Koolhaas et al., 2010), links with intra- or inter-specific affiliative behaviour have not been described to the best of the author’s knowledge. In addition, the low reliability of the Gregariousness towards Horses subscale (Chapter 3) would likely obscure any potential relationship of this factor with hair cortisol concentration.

Despite the lack of relationship between hair cortisol concentration and the other four personality factors, the model revealed a positive association between hair cortisol concentration and Agreeableness. This positive association was not expected on the basis of emerging results in the horse (Sauveroche et al., 2020), as Agreeableness and Dominance are generally considered two separate factors in animal personality structures (Gosling and John, 1999). However, the positive association between hair cortisol and Agreeableness described here is reminiscent of results in common Marmosets, in which hair cortisol is positively associated with the personality factor Sociability (Inoue-Murayama et al., 2018). In keeping with the terminology used in the EPT, Inoue-Murayama et al. (2018) note that this factor is conceptually close to two factors named “Agreeableness” and “Excitability” in a different Marmoset personality assessment (Iwanicki and Lehmann, 2015; Koski et al., 2017). A similar result is also reported in eastern chipmunks, with hair cortisol positively associated with “Docility”, a trait describing the ease of manipulation of an individual by human experimenters (Martin and Réale, 2008). The findings of this chapter therefore suggest that more agreeable horses, who are rated by caregivers as more obedient, willing, and well-mannered (Ijichi et al., 2013), experience higher levels of chronic HPA axis activity.

The positive association between Agreeableness and chronic HPA axis activity may be due to the fact that higher basal cortisol levels drive more agreeable behaviour in horses. Indeed, basal cortisol levels have been associated with the balance between punishment sensitivity and reward dependency in humans (Honk et al., 2003). Because horses are trained using negative reinforcement, the cues used in the context of equitation and daily management are associated with pressure or aversive stimuli (McGreevy and McLean, 2007). In addition, failure on the part of the horse to comply with human cues tends to be followed by an escalation of pressure intensity until the desired behaviour is offered (McGreevy and McLean, 2007); in the context of equestrian sports, it may also be followed by punishment (McLean and McGreevy, 2010). If horses with higher basal cortisol levels are more sensitive to punishment and to the pressures associated with negative reinforcement, this could result in more compliant behaviour and potentially better trainability. Indeed, higher cortisol levels have been associated with better performance in sports horses (Peeters et al., 2013).

However, it should be noted that the experimental design in this chapter made it possible to identify relationships between personality factors and hair cortisol concentration, but not causal links or their direction. It cannot be excluded that Agreeableness may instead drive higher cortisol levels, which can also reflect higher chronic stress levels (Russell et al., 2012). Passive or compliant behaviour in the horse is often thought to reflect the absence of a stress response, both within the industry and in the literature (Ijichi et al., 2013; Pearson et al., 2021). It is often assumed that horses that do not actively resist a potentially stressful procedure are not stressed by it and can carry on being exposed to it. However, a growing body of evidence shows that behavioural reactivity may not accurately reflect the intensity of stress responses, with compliant horses experiencing equivalent if not higher physiological stress responses than more behaviourally reactive horses (Yarnell, Hall and Billett, 2013; Squibb et al., 2018; Munsters et al., 2013b). Therefore, agreeable behaviour may result in more frequent and more prolonged exposure to stressors in the horse, leading to higher levels of chronic stress reflected by elevated hair cortisol. This result has important implications in terms of welfare, and further research is warranted to clarify the direction of causality between Agreeableness and elevated chronic HPA axis activity.

5.5 CONCLUSION

This chapter aimed to investigate chronic HPA axis activity, measured through hair cortisol concentration, as a potential physiological correlated of equine personality. Potential confounding factors of hair cortisol concentration were taken into account. In line with emerging results in the horse, sex, age, and hair cortisol were not found to impact on hair cortisol concentration. However, sampling location had an impact even within the mane, highlighting the importance of precisely standardising sampling location in future research. Neuroticism, Extraversion, Gregariousness towards People and Gregariousness towards Horses were not found to be related to hair cortisol concentration. The absence of link between Neuroticism and chronic HPA axis activity as measured through hair cortisol did not align with the *a priori* hypothesis. Taken together with results from the previous chapter, this suggests that Neuroticism may be independent of HPA axis function in the horse; this finding should be investigated further. However, a positive association was found between Agreeableness and hair cortisol concentration. This may be consistent with results linking basal cortisol levels with sensitivity to threat and punishment in humans. However, it is also possible that agreeable behaviour leads to higher chronic stress levels through increased exposure to stressors. This result has important implications in terms of welfare and further research is needed to explore the direction of causality between Agreeableness and elevated chronic HPA axis activity.

Chapter 6 Striatal dopamine as a potential correlate of equine personality

6.1 INTRODUCTION

Striatal dopamine levels have been identified as a driver of trait reward sensitivity in human (Depue and Collins, 1999) and rodent (de Boer, Buwalda and Koolhaas, 2017) models (reviewed in more detail in Section 1.3). More generally, dopamine functioning influences differences in incentive motivation and modalities of reinforcement learning (Maia and Frank, 2011; DeYoung, 2013). Therefore, traits such as susceptibility to stereotyies and cognitive style are correlates of personality in rodents (Coppens, de Boer and Koolhaas, 2010; de Boer, Buwalda and Koolhaas, 2017). Due to their highly conserved nature (O'Connell and Hofmann, 2012), dopaminergic networks are hypothesised to underlie the same traits in the horse (McBride et al., 2017). Empirically, differences in striatal dopamine levels have been shown to be associated with susceptibility to stereotypies in the horse (McBride and Hemmings, 2005). In addition, they are thought to underlie differences in learning style such as the tendency for habit formation and resistance to extinction (Hemmings, McBride and Hale, 2007; Roberts et al., 2015). These traits are highly relevant from an applied point of view, as they may impact on the trainability and sustainability of equines. However, at present, only sporadic genetic studies have explored a potential link between dopaminergic function and equine temperament (Momozawa et al., 2005b; Ninomiya et al., 2013). Dopaminergic function has therefore not yet clearly been established as a correlate of equine personality. Thus, at present little empirical evidence is available to evaluate the potential impact of selection on the basis of personality on other known outcomes of dopamine function, such as sensitivity to stereotypies or cognitive style.

Dopamine function can be monitored directly *in vivo* through the use of imaging techniques such as positron emission tomography (PET) scanning (Dang et al., 2017). Dopamine receptor densities can also be investigated post-mortem using homogenate-binding techniques (McBride and Hemmings, 2005). However, these techniques are invasive, labour intensive and costly, making them unsuitable for field studies of live subjects. As a result, dopamine function is routinely assessed through indirect markers (Dang et al., 2017). Converging evidence from pharmacological, epidemiological and cognitive studies suggests that baseline, or tonic, spontaneous blink rate (SBR) is positively related to tonic striatal dopaminergic activity (reviewed in: Jongkees and Colzato, 2016). Indeed, in rodents and non-human primates, tonic SBR increases following the administration of dopamine agonists, while it decreases following the administration of dopamine antagonists (Lawrence and Redmond, 1991; Kaminer et al., 2011; Elsworth et al., 1991). In addition, altered tonic SBRs are observed in atypical human populations with compromised striatal dopamine function: patients with Parkinson’s disease experience a loss of dopaminergic neurons in the striatum and have reduced SBR (Karson, 1983; Fitzpatrick et al., 2012), while patients with schizophrenia experience increased dopaminergic activity in the striatum and have increased SBR (Karson, 1983; Chen et al., 1996). Finally, SBR is related to dopamine-mediated modalities of reinforcement learning related with tonic levels of striatal dopamine, such as reward-driven behaviour (Slagter, Georgopoulou and Frank, 2015) and punishment aversion (Cavanagh et al., 2014). Some contradictory evidence is available; in particular, while tonic SBR has been clearly linked with striatal dopamine levels in atypical human populations, this link has not been replicated in healthy human subjects (Dang et al., 2017). Nevertheless, a wide body of evidence suggests baseline SBR can be used as a non-invasive indirect marker of tonic striatal dopamine levels in rodents, non-human primates and humans (Jongkees and Colzato, 2016).

At present, SBR has not been formally validated as an indicator of striatal dopamine function in horses through direct measurement or experimental manipulation of striatal dopamine. However, the brain networks involved are highly conserved across vertebrates (O'Connell and Hofmann, 2012; de Boer, Buwalda and Koolhaas, 2017) and findings relating to the dopaminergic influence on SBR have been shown to be generalisable between mammal species (Kaminer et al., 2011). Therefore, in recent years equine studies have used SBR as a proxy for dopamine function (Roberts et al., 2016, 2015), with findings that appear to mirror observations in humans (Roberts et al., 2015; Roebel and MacLean, 2007). This suggests that SBR may be a suitable indirect marker to non-invasively probe striatal dopamine function in the species.

Although the majority of research linking Extraversion to dopamine function in humans has been carried out using genetic or neural imaging techniques (for review see: DeYoung and Gray, 2009; Munafò, 2009), results coherent with this body of evidence have also been obtained using SBR as non-invasive indicator of dopamine function. To the author's knowledge, little evidence is available on the associations between SBR and the personality factors of the Five Factor Model. However, in keeping with theoretical frameworks (Depue and Collins, 1999), the only study available reports a weak but significant positive correlation between SBR and Extraversion (Unsworth, Robison and Miller, 2019). Links between SBR and factors in Eysenck's tridimensional personality model have been investigated in more depth, although the results available do not fully align with theoretical frameworks. A positive correlation between SBR and Extraversion was reported only in female subjects (Berenbaum and Williams, 1994). However, SBR was also found to be positively correlated with Neuroticism (Barbato et al., 2012) or with Psychoticism (Colzato et al., 2009), while Tharp and Pickering (2011) only report null results. These conflicting results have been attributed in part to differences in the psychometric tools used (Barbato et al., 2012; Jongkees and Colzato, 2016), as the Psychoticism scale used by Colzato et al. (2009) overlaps with facets of Extraversion (Barbato et al., 2012). In addition, differences in the methods used to record SBR may also contribute to the conflicting results observed (Jongkees and Colzato, 2016). Indeed, the only study to evidence a link between SBR and Extraversion measured blink rate while subjects viewed videos designed to elicit positive and negative affect rather than in primary gaze (Berenbaum and Williams, 1994, 1995). Interestingly, changes in SBR in response to affective stimuli have also been associated with scores on the Behavioural Activation System in Gray's BIS/BAS system (Berkovsky et al., 2019; Gros, 2011), which is thought to be closely related to Extraversion (DeYoung and Gray, 2009). Therefore, although methodological refinements are still necessary, SBR can successfully be used to probe the link between personality and dopamine levels in humans.

In continuity with the links reported in humans between SBR and personality, an emerging body of evidence suggests that SBR is also linked to equine personality. In horses, SBR has been shown to be positively correlated with "Anxiety" as measured in a 9-factor personality model (Roberts et al., 2016), and with EPT-Neuroticism (Loasby, 2018). SBR was also found to be negatively correlated with temperament trait "Docility" (Roberts et al., 2016), although the effect size was small (r = -0.21). Roberts et al. (2016) hypothesise that the links between SBR and Anxiety and Docility may be mediated by stress and reflect the differential impact of chronic, insurmountable stress on dopamine function in individuals with active (high Anxious) and passive (high Docile) coping styles. Therefore, it is not yet clear from emerging results whether SBR is linked to Extraversion-like equine personality factors, as is the case in humans. Nevertheless, these results do suggest that SBR could be used to explore the relationship between dopamine function and personality dimensions in the horse.

Phasic dopamine release in response to external stimuli may cause short-term variations in SBR and act as confounding factors when measuring tonic SBR (Jongkees and Colzato, 2016). In addition, factors such as eye dryness also results in phasic variations in SBR in humans (Al-Abdulmunem, 1999) and horses (Best et al., 2018; Cherry et al., 2020). In humans, SBR increases during social interactions such as conversation (Doughty, 2001), while it decreases with increasing mental workload (Lean and Shan, 2012) or in conditions requiring visual focus such as reading (Doughty, 2001) or attending to a moving stimulus (Bacher and Allen, 2009). In horses, phasic SBR responses have been documented during attentional processes and stress responses (Mott, Hawthorne and McBride, 2020; Merkies et al., 2019). Emerging results indicate that SBR decreases following exposure to an auditory startle (Mott, Hawthorne and McBride, 2018) and during focussed attention (Merkies et al., 2019), but may increase when exposure to a stressor is prolonged beyond the initial startle response (Mott, Hawthorne and McBride, 2020). Human blink characteristics, including SBR, are strongly affected by wakefulness state and fatigue: strong increases in SBR are observed during sleep deprivation, sleepiness or drowsiness (Cori et al., 2019). To date, little information is available on the impact of wakefulness state or fatigue on SBR in animals. SBR is therefore a highly sensitive measure and measures of tonic SBR are easily confounded by phasic responses to external stimuli.

To account for the impact of phasic variations on SBR, tonic SBR is assessed in primary gaze in humans, i.e. with subjects looking straight ahead to a neutral stimulus and not accomplishing any other mental or physical task concurrently (Doughty, 2001; Jongkees and Colzato, 2016). However, these conditions may be difficult to achieve when assessing SBR in animals. In particular, while the impact of attentional processes may be minimised by collecting blink rate data in a quiet, familiar environment and with the animal loosely restrained (Best et al., 2018; Roberts et al., 2016), any potential impact of wakefulness state would be much more difficult to control without interfering with the animals. Therefore, phasic changes in SBR relating to attentional processes or dozing behaviour should be controlled for or taken into account when assessing the relationship between tonic SBR and personality in animals.

Individual differences in dopamine function have therefore been linked with personality dimensions linked with reward sensitivity in humans (DeYoung and Gray, 2009) and rodent models (de Boer, Buwalda and Koolhaas, 2017). In the horse, while preliminary data is available to link reward sensitivity and striatal dopamine (Hemmings, McBride and Hale, 2007; Roberts et al., 2015), no links are currently documented between personality and dopamine function. The aim of this chapter was therefore to investigate whether striatal dopamine levels could be identified as a correlate of equine personality. To this end, SBR was measured at rest in a sample of 20 horses and compared with personality scores as measured using the EPT. The percentage of time spent dozing during the SBR observation period was included in the analysis to account for confounding variations in SBR linked with wakefulness state. In line with biological models of personality linking Extraversion with dopamine function (Depue and Collins, 1999), it was hypothesised that Extraversion and SBR would be positively related. In addition, on the basis of existing data in equines (Roberts et al., 2016; Loasby, 2018), it was hypothesised that Neuroticism and SBR would also be positively correlated. Finally, given that SBR increases with fatigue and drowsiness in human subjects (Cori et al., 2019), it was hypothesised that SBR would increase with the percentage of time spent dozing.

6.2 METHODS

6.2.1 HORSES AND MANAGEMENT

A total of N=20 horses were recruited from the Brackenhurst Equestrian Centre for inclusion in this study. This group of horses are a subsample of the population that has been used throughout this thesis. The sample comprised 13 geldings and 7 mares, with a mean age of 13.1 ± 3.75 years. 13 breeds were represented, with the most common being Irish sports horse (n=4), Connemara (n=3), Cob (n=2) and Thoroughbred (n=2). All horses had lived at Brackenhurst Equestrian Centre for a minimum of a year at the time of data collection. Data collection took place on 3 consecutive days in August 2019, during the early stages of the horses' summer conditioning program. During their summer conditioning program horses were kept on grass pasture in groups at night but brought in to the yard during the day. Housing was in individual stables (n=16) or combi barn in pairs (field shelter with small outdoor paddock within the yard: n=4). All horses were housed in their familiar home stables or barns. While on the yard they were fed hay or haylage, as well as supplementary feed according to individual requirements, in order to meet National Research Council nutritional guidelines (National Research Council, 2007). Water was available *ad lib* both at pasture and in stables. Members of the Brackenhurst Equestrian Centre staff gave the horses light exercise once daily, following a program aiming at progressively building fitness back up after the summer break. More details of sample demographics as well as management and exercise regimes during the summer conditioning program are available in Chapter 2.

6.2.2 BASELINE SPONTANEOUS BLINK RATE MEASUREMENTS

6.2.2.1 Video recordings

Thirty minutes long video recordings of undisturbed horses' face and eyes were obtained on three consecutive days. The impact of potential confounding factors such as time of day, exercise or feeding on spontaneous blink rate has not yet been established. Therefore, in order to minimise any impact of these potential confounding factors, horses were divided into 3 groups and the order in which the groups were filmed was rotated over the three days of data collection, so that a variety of conditions was represented for all horses. Footage was obtained using tripod-mounted GoPro Hero 7s (GoPro, San Mateo, California, US) set to record continuously at 1440 pixels/60 frames per second in the Wide FOV mode.

Horses were caught and loosely tethered in their stables using a headcollar and leadrope. The tripod-mounted GoPro was positioned inside the stable to obtain a clear view of the horse's face from the side and set to record. The horses were then left undisturbed for 30 minutes. The experimenters exited the stables but remained within sight to ensure horses remained in the shot and did not make contact with the equipment. Horses were only disturbed if it became necessary to move them back in the shot or to avoid a risk of damage to themselves or the equipment. After 30 minutes, the recording was stopped, the tripod taken out of the stable and the horse untethered. Video files were then transferred to a PC for analysis.

Videos were filmed from the side, with only one eye was visible in the footage for the majority of the time. Therefore, unilateral observations were carried out throughout. Due to the lateral position of the horse's eyes, this is the most frequently used method to determine blink rate in equine studies (Roberts et al., 2015; Merkies et al., 2019; Roberts et al., 2016). However, unlike in other equine studies (Roberts et al., 2015; Merkies et al., 2019; Roberts et al., 2016), the side of the eye observed was not standardised between horses. This was because light conditions differed between the horse's home stables and the decision was made to adapt camera placement to optimize exposure and footage quality. While unilateral blinks do occur in horses, their frequency does not differ significantly for the left and right eye (Best et al., 2018). In addition, Cherry et al. (2020) found that the blink rate measured unilaterally from the left and the right eye in the same horses did not significantly differ. Therefore, the side observed should not significantly impact the blink count and footage quality was prioritized.

6.2.2.2 Video processing and footage selection

Footage was deemed unsuitable for analysis if (1) the horse's eye was not visible in the shot, (2) the eye was visible but difficult to observe due to the horse's position, or (3) the horse was currently being disturbed by the experimenter in order to reposition them in the shot. All other footage, including sections featuring responses to normal yard activities, were considered suitable for analysis. Ahead of blink counts, the videos were processed using OpenShot Video Editor (OpenShot Studios, Rockwall, US). Where necessary, brightness and contrast were adjusted so that the eye was easily visible. In addition, all sections of footage selected for blink count were zoomed in and cropped to centre on the eye. This ensured that the eye was of a sufficient size to allow for reliable blink counts. The videos were not modified in any other way prior to blink count.

Blink counts were carried out in 5-minute segments of footage suitable for analysis (Best et al., 2018; Cherry et al., 2020), beginning at minute 10 of the original videos as it has been suggested a habituation period to recording conditions may be necessary for SBR to stabilise (Jongkees and Colzato, 2016). If all footage between minute 10 and minute 15 was suitable for analysis, then this was used as the 5-minute segment. However, if unsuitable footage was present between minutes 10 and 15, the segment retained for analysis was extended beyond the 15 minute mark for a duration equivalent to that of the unsuitable section, so that the resulting video contained a total of 5 minutes of footage suitable for analysis.

6.2.2.3 Blink count

A variety of eyelid movements associated with blinking have been reported in the horse (Wathan et al., 2015). Full blinks, defined in humans as "bilateral paroxysmal brief repetitive eye closures occurring continuously" (Karson, 1983), have been reported in horses (Roberts et al., 2016; Best et al., 2018; Merkies et al., 2019). However, partial blinks in which the full surface of the cornea is not covered by the upper eyelid have also been reported (Wathan et al., 2015; Best et al., 2018), as well as eyelid twitches (Merkies et al., 2019). To account for this variety in blinking behaviour, here a blink was defined as "any appreciable downward movement of either upper eyelid to cover some or all of the corneal surface, immediately followed by eye reopening".

The blink counts were carried out in BORIS Video Analysis Software (Friard and Gamba, 2016). Pilot studies showed that repeated blink counts were highly consistent when carried by the same experimenter, indicating that a single count could produce a reliable number of total blinks in the video (see Appendix B, Section A). Therefore, the total number of blinks was only counted once in each video; all videos were analysed by the author. To reproduce the conditions used for the intra-rater analysis as closely as possible, footage was viewed at speed 0.50x and blinks were recorded using the Behaviour Coding Pad placed in the same corner of the screen as the online tally counter.

In addition, pilot studies revealed that wakefulness state significantly impacts SBR (see Appendix A, Section B). Wakefulness state was therefore recorded throughout the videos so it could be controlled for in the analysis. Wakefulness state was coded as “Active”, “Dozing” or “Eyes closed”. “Eyes closed” was defined as a state in which the corneal surface was fully covered by the eyelids for more than a second (Best et al., 2018). This state was later excluded from the SBR analysis. Dozing was identified based on previously published ethograms as a behavioural state in which the horse stands inactive with the head lowered, the lower lip relaxed, and eyes partially closed; muscle twitches and leaning behaviour may be observed or weight may be borne on three legs (McDonnell, 2003). States were mutually exclusive and any behaviour that was not “Eye closed” or “Dozing” was recorded as “Active”. Blinks and wakefulness states were recorded during separate viewings.

For each video, SBR was calculated in blinks/minute by dividing the total number of blinks recorded by the duration of the video, excluding any “Eyes closed” segments. The percentage of observation time spent dozing was also calculated. It was defined as the total duration spent in the Dozing state divided by the duration of the video (excluding any “Eyes closed” segments), multiplied by 100.

6.2.3 PERSONALITY SCORES

Personality scores were obtained for all horses in the sample using the Equine Personality Test (Ijichi et al., 2013). This questionnaire uses subjective trait ratings by a familiar handler to assign a horse with a continuous score between 1 and 5 on five equine personality factors: Agreeableness, Neuroticism, Extraversion, Gregariousness towards People, and Gregariousness towards Horses. Here, ratings were provided for all horses by the same 3 primary caregivers, who had been familiar with the horses for a minimum of a year at the time of assessment. Questionnaires were scored for each rater individually, and for each horse the scores given by the 3 raters were then averaged to obtain the final set of personality scores used in the analysis. Further details on questionnaire items, scoring methods and validity and reliability of the Equine Personality Test are available in Chapter 3.

6.2.4 **STATISTICAL ANALYSIS**

All statistical analysis in this chapter was performed using R (version 3.6.1: R Core Team, 2019). To accommodate the significant impact of wakefulness state identified in pilot studies (see Appendix B, Section B), the percentage of time spent dozing in each video was included as a covariate in the final analysis to investigate links between SBR and personality factors. Given the very low consistency of SBR over the three days of data collection (see Appendix B, Section C), an average of the three repeated blink counts was not considered a meaningful measure. Instead, all counts were entered separately into the model. A Gaussian Generalized Additive Mixed Model (GAMM) was used to model SBR as a function of the 5 personality factors and the percentage of time spent dozing in each video, taking into account the dependency introduced by the repeated measures. Backward model selection was used to identify which variables to retain in order to obtain the best-fitting model. Figures were generated using the R package ***ggplot2*** from the ***tidyverse*** collection of packages (Wickham et al., 2019)

6.2.4.1 Data exploration

Data exploration was undertaken prior to model fitting, following the protocol suggested by Zuur, Ieno and Elphick (2010). There were no zeros, missing values or significant outliers in the response or explanatory variables. The response variable was continuous, strictly positive, homogenous, and normally distributed (Shapiro-Wilk test: W = 0.971, p = 0.171). There was no collinearity (all r ≤ 0.6) or multicollinearity (all VIF < 2.5) among explanatory variables. There was no graphical evidence of interactions between explanatory variables and, therefore, no interaction terms were included in the fitted model. Exploratory Spearman's correlations between SBR and personality factors were also conducted.

6.2.4.2 Model formulation

Given that the response variable was continuous, strictly positive, homogenous and normal, data were modelled with a Gaussian distribution with identity link function. There was dependency in the model due to the repeated measures taken from the same individual on different days. Consequently, horse identity was included in the model as a random term. In addition, the percentage of time spent dozing was found to be non-linearly related to SBR and this variable was fitted with a smoothing term. As a consequence, SBR was modelled as a function of the covariates using a Gaussian GAMM, which took the following form:

$$SBR_{ij} \sim N(\mu_{ij}, \sigma^2{}_{ij})$$

$$E(SBR_{ij}) = \mu_{ij}$$

$$\mu_{ij} = \beta_0 + f(Dozing_{ij}) + agree_{ij} + neur_{ij} + extr_{ij} + gregP_{ij} + gregH_{ij} + Horse_j$$

$$Horse_j \sim N(0, \sigma^2{}_{Horse})$$

Where SBR_{ij} was the spontaneous blink rate in observation *i* for horse *j*. The variables $agree_{ij}$, $neur_{ij}$, $extr_{ij}$, $gregP_{ij}$, and $gregH_{ij}$ were continuous covariates representing Agreeableness, Neuroticism, Extraversion, Gregariousness towards People, and Gregariousness towards Horses, respectively. $f(Dozing_{ij})$ was a smooth function to model changes in SBR as a non-linear function of the percentage of time spent dozing during the observation period. $Horse_j$ was included as a random intercept in the model to account for the repeated measures taken on different days for each individual and was assumed to be normally distributed, with a mean of zero and variance σ^2. The model was fitted using the *gamm4()* function from the ***gamm4*** package (Wood and Scheipl, 2017).

6.2.4.3 Model selection

Manual backward selection was used in order to identify the best-fitting plausible model. This procedure was performed through an iterative process using the Akaike Information Criterion (AIC) as measure of quality of fit of each model, with a lower AIC indicating a better fit. For each iteration, AIC was calculated for the full model and for all possible models obtained by removing a single covariate. For each iteration, the covariate whose removal resulted in the biggest improvement in the quality of fit was removed. The resulting model was then entered as the full model in the next iteration. Iterations were repeated until removing further variables resulted in a markedly higher AIC, i.e. a worse fitting model. The model obtained through manual backward selection is referred to thereafter as “the final model”.

6.2.4.4 Model validation

The fit of the final model was assessed through graphical means by plotting the residuals against both the fitted values and the retained covariates. There were no obvious patterns in the residuals and the fit of the model was considered satisfactory.

6.3 RESULTS

6.3.1 CORRELATIONS BETWEEN SBR AND PERSONALITY FACTORS HYPOTHESISED AS DOPAMINE-DRIVEN

There was a significant, positive correlation between SBR and Neuroticism (r=0.33, p=0.011). There was no significant correlation between SBR and Extraversion (r=-0.19, p=0.141).

6.3.2 MODEL SELECTION

Manual backward selection resulted in the successive removal of covariates $gregP_{ij}$, $neur_{ij}$, $extr_{ij}$ and $agree_{ij}$. However, covariate $gregH_{ij}$ (Gregariousness towards Horses) as well as the smoothed term $f(Dozing_{ij})$ (percentage of the observation period spent dozing) were retained in the best fitting model. The final model therefore took the form:

$$SBR_{ij} \sim N(\mu_{ij}, \sigma^2{}_{ij})$$

$$E(SBR_{ij}) = \mu_{ij}$$

$$\mu_{ij} = \beta_0 + f(Dozing_{ij}) + gregH_{ij} + Horse_j$$

$$Horse_j \sim N(0, \sigma^2{}_{Horse})$$

6.3.3 IMPACT OF COVARIATES ON SBR

There was a significant negative association between Gregariousness toward Horses and SBR (Table 6.1a: p = 0.002); more gregarious horses had lower SBR (Figure 6.1a). In addition, the smoothed term showed that the percentage of time spent dozing had a highly significant (Table 6.1b: p < 0.0001), non-linear effect on SBR. For low values of time spent dozing, SBR increased with the time spent dozing, up to a peak value for 25% of time spent dozing. SBR then decreased as time spent dozing increased, with a second smaller inflection around 80-85% of time spent dozing (Figure 6.1b).

Table 6.1 - Summary of a Gaussian GAMM to examine SBR as a function of ***a.*** *linear term Gregariousness towards Horses and* ***b.*** *smoothed term Percentage of time spent dozing. Samples from different horses were fitted as random intercepts.* $N_{obs} = 60$.

a.	**Estimate**	**Std. Error**	**t**	**P value**
(Intercept)	23.39	3.97	5.89	<0.001
Gregariousness towards Horses	-3.79	1.13	-3.34	0.002

b.	**Est. degrees of freedom**	**F**	**P value**
Percentage of time spent dozing	4.88	8.72	<0.001

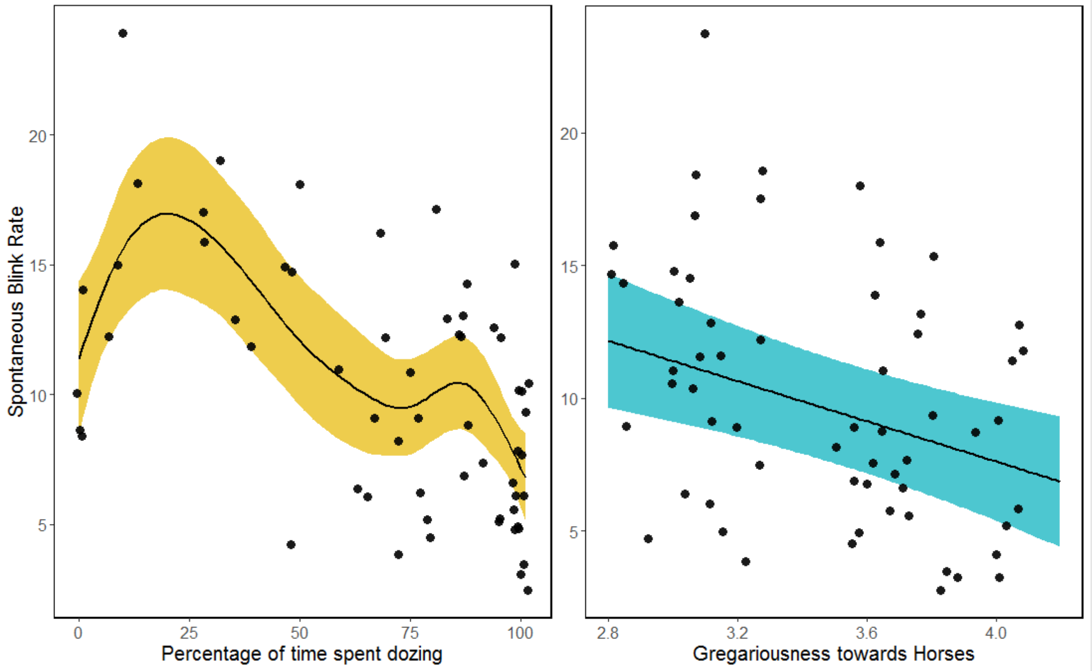

Figure 6.1 - Fitted values of SBR against ***a.*** *Percentage of time spent dozing, and* ***b.*** *Gregariousness towards Horses, modelled using a Gaussian GAMM. Coloured bands indicate 95% confidence intervals around the fitted line. Black circles are observed values for SBR.*

6.4 DISCUSSION

Individual differences in tonic striatal dopamine levels have been proposed as a driver of personality factors such as Extraversion in humans. In horses, although similar links have been suggested, little is known about the relationship between the personality factors measured by the Equine Personality Test and dopamine. The present study investigated the relationship between personality scores in the EPT and baseline SBR, an indirect marker of striatal dopamine function, in a sample of riding school horses (n=20). The impact of dozing behaviour on SBR was also taken into account. The percentage of time spent dozing during the observation period was found to strongly and non-linearly impact SBR. Contrary to hypotheses, neither Neuroticism nor Extraversion contributed to explaining SBR. However, an unexpected negative association between SBR and personality factor Gregariousness towards Horses was found.

Dozing was included as a covariate in the analysis following pilot observations suggesting SBR was strongly reduced while dozing compared to an active state. Subsequently, the percentage of time spent dozing during the observation period was found to have a strong, non-linear impact on SBR. SBR increased with dozing up to a peak at 25% of time spent dozing, then decreased as dozing increased further, with a second, smaller peak around 85% of time spent dozing. This strong impact of dozing behaviour on SBR was unexpected as it has not yet been described in the equine literature. It is possible that the positive relationship observed between SBR and low values of percentage of time spent dozing reflected the impact of focussed attention. Horses who remained alert during the entire observation period may have expressed more focussed visual attention, which is associated with a reduction of SBR in horses (Mott, Hawthorne and McBride, 2020; Merkies et al., 2019). However, the reduction in SBR associated with high levels of dozing was unexpected as findings in humans repeatedly show that fatigue and drowsiness result in an increase in blink rate (Cori et al., 2019; Jongkees and Colzato, 2016). However, it should be noted that these states may not be comparable to dozing behaviour in horses. In human subjects, fatigue and drowsiness are characterised by efforts to remain in a wakeful state, which is promoted by dopamine (Monti and Jantos, 2008; Oishi and Lazarus, 2017). By contrast, dozing in horses corresponds to an early stage of sleep. In addition, while the increase in SBR during drowsiness in humans appears dopamine-mediated, it should be noted that factors other than dopamine levels contribute to regulating blink rate (Kaminer et al., 2011). In particular, blink behaviour has been linked with tear film replacement in humans (Al-Abdulmunem, 1999), rodents (Kaminer et al., 2011) and horses (Best et al., 2018; Cherry et al., 2020). Air flow over the ocular surface may contribute to tear film break up and is associated with increased blink rate (Wu et al., 2014). In horses, it has been suggested that increased movement may result in increases in air flow over the ocular surface, and therefore in increases in SBR (Cherry et al., 2020). Dozing behaviour in horses is characterised by relaxation, a lack of movement and partially closed eyes (McDonnell, 2003); these factors may contribute to minimise tear film break-up and the need to blink, resulting in a decrease in SBR. To the author’s knowledge, this study is the first to demonstrate an impact of dozing behaviour on SBR in horses. More research is needed to fully characterise the dual impact of short-term attentional processes and dozing behaviour on SBR in the horse, in order to inform protocols used to measure tonic SBR. Given the large proportion of time horses in the sample spent dozing, it appears likely that the values of SBR obtained in other equine studies, in which horses were observed while tethered for as long as 30 minutes (e.g. Roberts et al., 2015, 2016), would have been affected by dozing.

The absence of link between SBR and Extraversion is in line with previous findings in the horse (Loasby, 2018; Roberts et al., 2016), although it is unexpected from a comparative point of view (Depue and Collins, 1999). In line with the results presented here, no links between SBR and EPT-Extraversion has been reported in a previous study in a sub-sample of the horses used in this thesis (Loasby, 2018). In addition, Roberts et al. (2016) did not report a correlation between SBR and Excitability, the factor in their equine personality model most similar to Extraversion at the trait level. Therefore, this result appears to strengthen existing evidence that does not link Extraversion with SBR in the horse, thus implying that equine Extraversion may not be linked with dopamine levels. Nevertheless, this is highly unexpected from a comparative point of view, given the strong evidence linking trait reward sensitivity to dopamine levels (Depue and Collins, 1999; Munafò, 2009; DeYoung and Gray, 2009; de Boer, Buwalda and Koolhaas, 2017) and the highly conserved character of the neural networks involved (O'Connell and Hofmann, 2012).

A possible explanation for this contradictory finding in the horse may be that EPT-Extraversion is conceptually different from FFM-Extraversion and trait sensitivity in rodents, and may therefore have different underlying neurophysiological drivers. However, in addition to similarities to FFM-Extraversion at the trait level (see Section 1.4.3), EPT-Extraversion has been shown to reflect differences in coping styles (Ijichi et al., 2013). Another potential explanation for this finding may be that tonic SBR is not the most salient indicator of dopamine levels to use when exploring links with Extraversion. Despite the strong evidence linking Extraversion with dopaminergic function in humans (DeYoung and Gray, 2009), the studies investigating links between tonic SBR and human personality report contradictory results (Barbato et al., 2012; Colzato et al., 2009). By contrast, phasic SBR responses to standardised affective stimuli are correlated with Extraversion and similar traits linked with reward sensitivity (Berenbaum and Williams, 1994; Berkovsky et al., 2019). A future avenue for research may be to explore the relationship between personality and phasic SBR responses to emotive situations in horses. In addition, despite time and logistic constraints, cognitive studies may provide a clearer insight into the relationship between Extraversion and dopaminergic function.

There was a significant positive correlation between SBR and Neuroticism, in accordance with previous findings in humans and equines showing a positive correlation between SBR and Eysenck's Neuroticism (Barbato et al., 2012), EPT-Neuroticism (Loasby, 2018) or "Anxiety" (Roberts et al., 2016). However, Neuroticism was not retained as an explanatory variable of SBR in the modelling analysis. This contradictory finding might in part be explained by the inclusion of the covariate "percentage of time spent dozing" in the analysis, as this covariate appears to have captured phasic changes in SBR in response to wakefulness state and possibly attentional focus on the environment. It could be hypothesised that the observed impact of Neuroticism on SBR is in fact mediated by short-term behaviour or responses during the observation period. For instance, more stress sensitive horses may have spent less time dozing during the observation period, resulting in a higher recorded SBR. However, it should be noted that the percentage of time spent dozing did not significantly correlate with any of the personality factors, weakening this hypothesis. In addition, exposure to an environmental stressor has been shown to result in an increase in blink rate in the horse following the initial startle response (Mott, Hawthorne and McBride, 2020) and blink rate appears sensitive to even very mild stressors. More stress sensitive horses may have shown short term stress responses to environmental stressors more often during the observation period, resulting in higher SBR. This suggests that more research is needed to clarify the relationship between trait stress-sensitivity, short-term stress responses and SBR in the horse, in order to inform hypotheses linking Neuroticism to central dopamine levels (Roberts et al., 2016).

Finally, there was a significant negative impact of Gregariousness towards Horses on SBR, with this covariate being the only personality factor retained as an explanatory variable of SBR in the modelling analysis. This finding implies that more gregarious horses may have lower levels of tonic striatal dopamine. This result was unexpected as such an association between SBR and sociability towards conspecifics has not previously been reported in the horse (Loasby, 2018; Roberts et al., 2016). In addition, while dopamine has been identified as a potential driver of sociability through the rewarding effect of social contact (Depue and Morrone-Strupinsky, 2005), the direction of the effect reported here is in contradiction with both theoretical models of underlying drivers of affiliation (Depue and Collins, 1999; Depue and Morrone-Strupinsky, 2005), and empirical results that suggest a positive link between sociability and dopamine levels (Adamczyk et al., 2012; Shimohata et al., 2017; Bariselli et al., 2018).

A potential explanation for this unexpected finding is that horses were observed while individually stabled and were therefore in a situation of relative social isolation. In the horse, short term social isolation has been shown to result in at least a transient decrease in blink rate (Merkies et al., 2019). Although horses were not removed from their normal management for the purpose of the study and are assumed to be habituated to those conditions, it could be hypothesised that more gregarious horses were more affected by the relative social isolation resulting from single-housing conditions, leading to the observed reduction in SBR. However, the important limitations associated with the Gregariousness towards Horses scale of the EPT (Chapter 3) may affect the reliability of this result. While the predictive validity of other scales of EPT has been explored (Ijichi et al., 2013), the predictive validity of the Gregariousness towards Horses for social behaviour has not yet been established. In addition, this scale has been shown to have limited inter-rater and test-retest reliability (Chapter 3). Therefore, more work is needed to confirm the above finding and assess its applied relevance, using a valid and reliable measure of social behaviour and trait-affiliation/sociability.

6.5 CONCLUSION

The aim of this chapter was to explore the relationship between the five equine personality factors measured by the EPT and individual variations in striatal dopamine levels, measured indirectly through SBR. This chapter's findings reveal important methodological issues relating to the collection of SBR data in the horse. Wakefulness state was shown to exert a very strong, non-linear phasic impact on SBR. This impact had not yet been described in the horse. It is hypothesised that the non-linear pattern observed here may represent the dual impact of attentional processes and dozing behaviour on SBR. This chapter's results suggest that SBR measured in tethered horses in familiar environments may be strongly influenced by phasic responses rather than representing a true measure of tonic SBR, and therefore tonic striatal dopamine. More research is needed to clarify the impact of attentional processes and wakefulness state on SBR, and to establish a protocol enabling repeatable measurements of true tonic SBR in the horse. This chapter's findings on the relationship between equine personality factors and SBR contradict the hypotheses proposed. No relationship between SBR and Extraversion or Neuroticism was observed, suggesting that there may be no relationship between individual variations in striatal dopamine levels and those personality factors as measured by the EPT. By contrast, an unexpected negative association between SBR and Gregariousness towards Horses, a personality factor that was not thought to be underpinned by the dopaminergic function, was reported. Future work should focus on replicating this unexpected finding and further probing the relationship between equine personality factors and individual variations in striatal dopamine levels. Given the methodological issues raised here around SBR, this second aim may be more reliably achieved by using SBR in combination with other indirect markers of striatal dopamine, such as cognitive characteristics.

Chapter 7 General Discussion

Personality is identified by equestrians as one of the most relevant characteristics of a horse with regards to its performance and suitability for a role (Graf, König von Borstel and Gauly, 2013). A consistent profile of desired personality emerges from surveys of equestrians at all levels, highlighting a desire for a compliant horse with limited behavioural reactivity to novelty or environmental challenges (Suwała et al., 2016; Górecka-Bruzda et al., 2011; Graf, König von Borstel and Gauly, 2013). Equine personality research in the last few decades has focussed on the development of assessment tools that may be used either in the field or in research (Rankins and Wickens, 2020). As a result, objective tests now enable selection on the basis of personality for breeding (Graf, König von Borstel and Gauly, 2014) or for more demanding roles such as a mounted police horse (Pierard, McGreevy and Geers, 2017) or sports horse (Lansade et al., 2016). From a fundamental research point of view, the development of multiple subjective trait-rating assessment tools (e.g. Momozawa et al., 2005a; Lloyd et al., 2007; McGrogan, Hutchison and King, 2008; Ijichi et al., 2013) has given rise to a tentative model of equine personality (reviewed in Section 1.4.3). This emerging model is coherent with comparative personality studies that indicate some personality traits are conserved across multiple taxa (Gosling and John, 1999; Gosling, 2001). However, to date, research on correlates of equine personality remains relatively limited beyond the behavioural level (Rankins and Wickens, 2020). In particular, little research has systematically addressed the question of the biological basis and physiological correlates of personality in the horse (Rankins and Wickens, 2020).

In other species (including humans), physiological correlates of personality have been identified at both the central (DeYoung and Gray, 2009) and peripheral level (Ormel et al., 2013; Chida and Hamer, 2008). At the central level, differences in serotonin and dopamine function have been identified as drivers of personality factors such as Neuroticism and Extraversion (DeYoung and Gray, 2009), and of coping style in animals (de Boer, Buwalda and Koolhaas, 2017). In addition, at the peripheral level, the reactivity of physiological systems linked with the adaptive stress response has been associated with Neuroticism (Chida and Hamer, 2008; Ormel et al., 2013) and coping styles (Koolhaas et al., 2010). This is crucial as these physiological correlates then go on to mediate links between personality and outcomes such as cognitive style (Coppens, de Boer and Koolhaas, 2010) and susceptibility to stress related diseases (Koolhaas, 2008). Similar links between neurophysiological parameters and personality have only very tentatively been identified in the horse to date (Rankins and Wickens, 2020). Therefore, at present, it is unknown whether a particular physiological profile is being selected alongside the desired equine personality profile, and if it is, whether this physiological profile is advantageous. This thesis aimed to address this gap by exploring the relationship between equine personality dimensions measured using a species-specific trait-based questionnaire (Ijichi et al., 2013) and neurophysiological correlates of personality identified in other species.

7.1 SIGNIFICANCE AND LIMITATIONS OF FINDINGS

7.1.1 VALIDITY AND RELIABILITY OF THE EPT

The Equine Personality Test (EPT; Ijichi et al., 2013) was chosen as the equine personality assessment tool for this research. This choice was motivated by the use of a psychometric approach in developing the questionnaire (Creighton, personal communication; Ijichi et al., 2013), as well as its integration within comparative models of personality such as the conserved Five Factor Model (Gosling and John, 1999) and the coping style framework

(Koolhaas et al., 1999; Ijichi, 2014). The concurrent and predictive validity of the EPT has previously been demonstrated: Neuroticism and Extraversion scores predict behavioural responses related to those traits in standardised tests (Ijichi et al., 2013), and the expression of pain-related behaviour (Ijichi, Collins and Elwood, 2014). However, some elements of validation of the EPT had yet to be carried out to ensure the EPT yielded valid and reliable personality measures. In particular, the EPT had yet to be tested for internal consistency of the factors (Simms and Watson, 2007), inter-rater reliability (Gosling and Vazire, 2002), and test-retest reliability (Dingemanse and Wright, 2020). Chapter 3 therefore sought to evaluate whether the EPT met these criteria. Due to the factor structure of the EPT, each subscale was considered separately.

This investigation showed that the subscales measuring Agreeableness, Neuroticism, Extraversion and Gregariousness towards People showed very good internal consistency, inter-rater reliability and test-retest reliability. These findings are not surprising as the EPT was developed using psychometric techniques recommended for the construction of personality assessment scales (Ijichi et al., 2013; Creighton, personnal communication). Taken together with the results from Ijichi et al. (2013, 2014) demonstrating predictive validity for two of the subscales, this work contributes to making the EPT the only subjective equine personality questionnaire to have been checked against all four criteria of a valid and reliable personality assessment tool laid out by Gosling & Vazire (2002) and Dingemanse and Wright (2020). Currently, the questionnaires developed by Momozawa et al. (2005) and Lloyd et al. (2007) are most commonly used by equine researchers (reviewed in Rankins & Wickens, 2020). However, the EPT now benefits from a more thorough validation than both these questionnaires. In particular, neither has been evaluated for test-retest reliability, a crucial aspect of personality assessment (Dingemanse and Wright, 2020). In addition, the results from Chapter 3 show that the EPT provides a particularly robust measure of the three factors Agreeableness, Neuroticism and Extraversion. This is relevant from an applied point of view, as these are the main factors of interest when selecting horses to fit the profile of personality desirable to equestrians. It is also relevant from a research point of view, as these are the three main factors identified as conserved across species (Gosling and John, 1999) and consistently identified by in equine personality models (reviewed in Section 1.4.3). Therefore, the results presented in Chapter 3 are novel and contribute to position the EPT as a highly relevant equine personality assessment tool, both for applied and research contexts. They also strengthen the credibility of the personality assessment used in subsequent chapters of this work.

One limitation of the findings in this chapter is that the high levels of inter-rater reliability observed for the Agreeableness, Neuroticism, Extraversion and Gregariousness towards People scales may not be generalisable to any sample of raters. Indeed, all raters used in this study were recruited from the same workplace and were knowledgeable not only about horse management but also equine science. This likely had a positive influence on their ability to accurately interpret everyday behaviour they observed in the horses (Bell et al., 2019), leading to increased agreement between raters on the usual responses of a given horse (Funder, 1995). For this reason, it is unclear from the results of this work whether the high inter-rater reliability observed here can be generalised to less specialised raters, such as amateur horse owners, who are more likely to misinterpret behavioural responses (Bell et al., 2019). In addition, traits in the EPT are presented as pairs of adjectives or descriptors without further definition, leaving them open to interpretation by the raters. It has previously been shown that equestrians vary in their definitions of terms commonly used to describe horse personality (Mills, 1998). Although the raters used in this study did not discuss their assessments of the target horses' personality when filling in the questionnaires, it is likely that they would have developed a shared vocabulary over the course of their daily professional interactions. This may have made them more likely to interpret the traits in the EPT similarly. Therefore, while the findings of this work can be used to support the reliability of the EPT when the assessment is carried out by equine professionals, additional work may be needed to ensure that the high level of inter-rater reliability described here is conserved when the assessment is carried out by a more varied group of raters.

Chapter 3 also revealed that not all subscales of the EPT satisfied the criteria laid out by Gosling & Vazire (2002) and Dingemanse and Wright (2020). Indeed, the Gregariousness towards Horses scale yielded poor internal consistency, poor inter-rater reliability and poor test-retest reliability. In addition, the concurrent validity of the Gregariousness towards Horses scale was not investigated by Ijichi et al. (2013). Therefore, in addition to the reliability issues highlighted by the present work, the Gregariousness toward Horses scale has not yet been shown to successfully predict social behaviour and trait-affiliation/sociability. Therefore, at present this scale does not provide an acceptable measure of the target personality factor. While this scale was included in the prospective studies carried out in the last two chapters of this work (Chapter 5, Chapter 6), this was strictly in order to inform potential avenues of future research, and results involving this scale should be interpreted with much caution.

A clear avenue for future research emerging from Chapter 3 is to revise the EPT with a focus on the Gregariousness towards Horses scale, in order to ensure all personality dimensions captured by the questionnaire are measured in a valid and reliable way. Such revisions are common in human (e.g. Eysenck et al., 1985; Mccrae & Costa, 2004) and animal (Ley, Bennett and Coleman, 2009; Olsen and Klemetsdal, 2017) personality questionnaires when problematic elements are identified. They are usually carried out by either replacing or adding additional items to the problematic scale, using psychometric methods as outlined by e.g. Simms & Watson (2007) to identify items of interest. Due to the numerous issues associated with the scale, one potential solution might be to disregard it altogether. However, trait-sociability emerges as a separate factor in most equine personality models (e.g.

Momozawa et al., 2005; Lloyd et al., 2007; Roberts et al., 2016), and simply removing this scale risks failing to capture a relevant dimension of equine personality. Revisions to the Gregariousness towards Horses scale, following psychometric guidelines (Simms and Watson, 2007), would therefore be more advisable. All items on the scale showed limited reliability; therefore, substituting those items for more reliable alternatives may prove fruitful. The traits "playful", "sociable", "popular", "friendliness – horses", and "competitiveness" have been shown to be rated reliably and to contribute to personality factors that appear conceptually close to Gregariousness towards Horses: "Sociability" (Lloyd et al., 2007) and "Horse-horse interaction" (Roberts et al., 2016). Therefore, these traits may constitute a relevant item pool for a revised Gregariousness towards Horses scale, subject to confirmation through factor analysis.

Once designed, the revised Gregariousness towards Horses scale should be tested for concurrent validity. One difficulty in achieving this is the lack of well-established method to evaluate gregariousness or sociability in the horse. While objective behavioural tests have been developed to evaluate the reactivity to isolation from conspecifics (Lansade, Bouissou and Erhard, 2007), it appears likely that behavioural expression in those tests may be influenced not only by sociability but also by Neuroticism and Extraversion, given the stressful nature of the challenge. Other tests of sociability trialled that lacked this stressful element, such as a social attraction test, did not yield consistent results over time and may therefore not be appropriate to evaluate a personality trait (Lansade, Bouissou and Erhard, 2007).

Once a revised version of the EPT satisfying Gosling & Vazire's (2002) criteria for all its scales is available, it should be used to confirm and investigate in more depth the findings of the present work. In particular, future work should aim to investigate the findings from Chapter 6 in more depth. Although this chapter failed to evidence hypothesised links between spontaneous blink rate (SBR) and Extraversion and Neuroticism, it revealed an unexpected negative association between SBR and Gregariousness towards Horses. The biological significance of this finding is not immediately obvious, as sociability is typically positively, rather than negatively, associated with dopamine levels in other species (Depue and Morrone-Strupinsky, 2005). While a reduction in SBR due to the impact of social isolation (Merkies et al., 2019) has been very tentatively proposed to explain the findings in Chapter 6, it seems highly likely that the reliability of this finding is limited, due to the shortcomings of the Gregariousness towards Horses identified in Chapter 3. Therefore, before more work is undertaken to better understand the significance of this result, it should be confirmed using a valid and reliable measure of trait-sociability in the domestic horse.

7.1.2 PHYSIOLOGICAL PROFILE OF PERSONALITY TYPES

The overarching aim of this thesis was to determine whether a physiological profile of personality types could be established in the horse, in line with findings in human (Ormel et al., 2013; DeYoung and Gray, 2009; Depue and Collins, 1999) and non-human (de Boer, Buwalda and Koolhaas, 2017) animals. Potential neurophysiological correlates chosen for this preliminary investigation were those that had been identified as correlates of personality in human and/or animal models, were of applied relevance to equine welfare or performance, and could be studied using strictly non-invasive methods. Therefore, at the peripheral level, cardiac, parasympathetic and HPA axis reactivity to stressors were evaluated (Chapter 4), as well as chronic HPA axis activity (Chapter 5). At the central level, striatal dopamine was considered (Chapter 6).

Based on findings in humans (reviewed in Ormel et al., 2013), links were suggested between Neuroticism and cardiac, parasympathetic and HPA axis reactivity to stressors, as well as baseline HPA axis activity. However, these physiological parameters have also been linked with coping style in the coping style framework (Koolhaas et al., 2010). Given that EPT-Extraversion is hypothesised to represent coping style in the horse (Ijichi et al., 2013), links with Extraversion were also tentatively proposed. Finally, given the strong evidence for a dopamine basis to Extraversion-like factors in human and non-human models (Depue and Collins, 1999; DeYoung and Gray, 2009; de Boer, Buwalda and Koolhaas, 2017), a positive link was hypothesised between Extraversion and striatal dopamine levels. Given the importance of compliant behaviour to equestrians (Górecka-Bruzda et al., 2011; Suwała et al., 2016), Agreeableness was also included in all analyses. However, it was not hypothesised to be linked with any of the potential physiological correlates considered here, as Agreeableness-like factors are thought to be underpinned by serotoninergic networks and largely independent of stress reactivity (de Boer, Buwalda and Koolhaas, 2017; DeYoung and Gray, 2009).

None of the hypothesised links between equine personality and physiological parameters were supported by the data collected in this project. Contrary to hypotheses, Chapter 4 revealed that the heart rate, heart rate variability and salivary cortisol responses to three experimental stressors were unrelated to Neuroticism and Extraversion. In addition, Chapter 5 showed that hair cortisol, a non-invasive indicator of chronic HPA axis activity, was not linked to either Neuroticism or Extraversion. Finally, Chapter 6 showed that spontaneous eye blink rate, a non-invasive proxy measure of tonic striatal dopamine activity, was not related with EPT measures of Extraversion. Therefore, no evidence was found in the data to support any of the proposed relationships between personality as measured by the EPT and potential neurophysiological correlates, either at the central or peripheral level. No physiological profile of equine personality types in continuity with those established in humans and non-human models can therefore be proposed on the basis of this work. Given the highly conserved character of the biological systems and personality traits considered (Gosling and John, 1999; Gosling, 2001), this is a very unexpected finding.

Methodological limitations may have contributed to the lack of findings in line with hypotheses. First, it could be suggested that the personality scores used in those analyses did not accurately reflect the personality dimensions they aim to measure. However, the work carried out in Chapter 3, in conjunction with the validation work carried out by Ijichi et al. (2013), strongly suggests that the personality scores used in Chapter 4, Chapter 5 and Chapter 6 were valid and reliable. Therefore, this criticism is unlikely to apply here. However, limitations of the non-invasive techniques used to measure autonomic and HPA axis reactivity and tonic striatal dopamine activity may have contributed to the null findings reported in these chapters (identified and discussed in depth in Chapter 4 and Chapter 6). Limitations were linked with the impact of potential confounding factors, such as attentional processes and short-term stress responses for spontaneous blink rate (Cherry et al., 2020; Merkies et al., 2019: discussed in more detail in Section 6.4). While the equipment used in this work, such as Polar heart rate monitors and ELISA assays, is standard for equine studies (e.g. Stucke et al., 2015; Sauer et al., 2019), their reliability is being increasingly questioned (e.g. Parker et al., 2009; Sauer et al., 2020: discussed in more detail in Section 4.4). Taken together, these methodological limitations could have impacted the reliability with which rank order or subtle individual variations were captured within the cohort. By contrast, studies in rodents and other laboratory animal models frequently employ invasive techniques that provide a much more direct measure of the physiological variable of interest and are less susceptible to the impact of confounding factors (e.g. plasma cortisol and adrenaline/noradrenaline concentrations: e.g. Koolhaas et al., 2010; experimental administration of dopamine agonists/antagonists: e.g. Benus et al., 1991). This may increase their ability to detect subtle individual differences. While the use of invasive techniques cannot be proposed in the domestic horse for ethical reasons, refinements to the protocols and equipment used here may help determine the extent of the impact of methodological limitations on the results presented in this work.

Characteristics of the sample used in the current study may also contribute to explain why the findings of this work do not align with models in other species. First, sample size must be taken into account as a potential limitation of this study. A sample size calculation based on the strength of previously published correlations between physiological parameters and personality factors in the horse (Momozawa *et al.*, 2003; Roberts *et al.*, 2016; Sauveroche *et al.*, 2020) revealed that a larger sample size may be needed to detect those effects with acceptable statistical power (see Chapter 2). The study may therefore have been underpowered, which may have limited its ability to document effects in line with those described in other species.

To ensure applied relevance, the work presented in this thesis was conducted on a diverse sample of horses chosen to be representative of the UK's leisure horse population in as far as that is possible (Hotchkiss, Reid and Christley, 2007; Hockenhull and Creighton, 2013) and kept under an industry-standard management system. This is in contrast to studies in rodent models, in which selected strains representing extremes of the personality trait of interest tend to be used (e.g. Koolhaas et al., 1999). These studies often yield much more clear-cut results when compared to studies using normal populations, in which the full range of the trait can be observed (Réale et al., 2007; Ferrari et al., 2013). Variations in previous life experience and training in the sample of horses likely acted as an additional source of individual differences that is usually not present in samples of laboratory or farm animals, whose life experience and management is generally standardised. Despite the use of a broadly standardised management regime at Brackenhurst Equestrian Centre, management is often tailored to horses' individual needs to ensure optimal welfare for each individual. Additional safeguards are often put in place for horses that are identified as having extreme personalities, especially highly neurotic horses, to avoid causing them distress. It is possible that this individual tailoring of management and use might have contributed to mask individual differences in physiology, as highly stress sensitive individuals may experience less stressful situations than the rest of the group. Although the use of a continuous population kept under industry standard management may contribute to explain the lack of continuity between the findings of this work and those reported in other species, it is considered to be a strength rather than a limitation of the current work, as it guarantees the applied relevance of the findings.

The very limited links between physiology and personality documented in this work contradict emerging findings in the horse published by other groups over the course of this project. For instance, Roberts et al. (2016) report correlations between spontaneous eye blink rate and personality factors “Anxiety” and “Docility”, as measured by the subjective personality questionnaire developed in the same study. In addition, Sauveroche et al. (2020) document correlations between hair cortisol and personality factors “Dominance”, “Anxiousness”, and “Excitability” as measured by Lloyd et al.'s Horse Personality Questionnaire (2007). It should be noted that both these studies used simple correlations in their analyses, meaning that they might have failed to account for confounding factors that were accounted for in the GLMMs used in this work. In addition, the personality assessment tools used in these studies have not been fully checked against the criteria of validity and reliability laid out by Gosling and Vazire (2002) and Dingemanse and Wright (2020) Nonetheless, their results appear better aligned with the literature in other species than those reported in the present work. Contradictory results on the links between personality and physiological variables have also been published in the human literature (e.g. Neuroticism and cortisol levels: reviewed in Ormel et al., 2013; Extraversion and spontaneous blink rate: Barbato et al., 2012). These discrepancies in findings are commonly attributed to differences in methodology, including personality assessment tool (Barbato et al., 2012), and method or endpoints used to collect physiological data (Ormel et al., 2013; Jongkees and Colzato, 2016). Differences in methodology may also help explain differences between the findings of the present thesis and those of Roberts et al (2016) and Sauveroche et al. (2020), and should be addressed by future research.

To ensure results from different studies carried out in the horse are comparable, future research should aim to document what links, if any, exist between the different equine personality questionnaires frequently used by researchers (Momozawa et al., 2005a; Lloyd et al., 2007; Ijichi et al., 2013). In addition, existing evidence should be carefully reviewed to develop standardised methodologies to measure physiological correlates, taking into account any confounding factors. This is particularly true for emerging measures such as hair cortisol concentration (Chapter 5) or spontaneous blink rate (Chapter 6). Chapter 5 revealed that sampling location acted as a confounding factor of hair cortisol concentration, consistent with previous findings in the horse (Banse et al., 2020). Both the current work and that by Sauveroche et al. (2020) document relationships between personality and hair cortisol concentration sampled from the mane. It is therefore recommended that future works use the same sampling site. In addition, Chapter 6 showed that wakefulness state exerts a very strong phasic influence on spontaneous blink rate, therefore confounding the putative relationship between this parameter and tonic dopamine function (Roberts et al., 2016). This finding adds to the current understanding of factors known to affect spontaneous blink rate in the horse, including attentional focus (Cherry et al., 2020) and acute stress responses (Mott, Hawthorne and McBride, 2020; Merkies et al., 2019). At present, it seems likely that the methods used to measure tonic spontaneous blink rate in the horse (e.g. Roberts et al., 2016, 2017) could be affected by some or all of these sources of phasic variations. Therefore, it is crucial that standardised measurement methods for tonic spontaneous blink rate is developed in the horse, if this parameter is to be used as a proxy measure of tonic dopamine. Resolving these methodological issues should help strengthen the conclusions drawn by future works, and ensure comparability between studies.

7.1.3 PERSONALITY AND WELFARE

Although the hypothesised neurophysiological profile of personality types did not emerge from this thesis, some correlations between personality and physiological parameters were discovered. In particular, results from Chapter 4 and Chapter 5 are relevant from a welfare point of view and merit further discussion. When taken together, the results of these two chapters echo other works which have expressed concern about the welfare of compliant, less behaviourally expressive horses (Yarnell, Hall and Billett, 2013; Squibb et al., 2018; Munsters et al., 2013b).

Chapter 4 identified no correlations between personality factors scored using the EPT and physiological reactivity to stressors. As discussed previously (see Section 4.4), this null finding may be due to the limited aversiveness of the stressors used for this sample of horses, and to limitations in the reliability of the methods used to monitor physiological responses. However, at present this result suggests that the EPT does not have predictive validity for physiological responses to stressors. This is a highly relevant finding from an applied point of view as the main application of equine personality assessment is to pair horses with appropriate homes, owners, or roles based on their individual characteristics (König von Borstel, 2013). While behavioural reactivity and compliance are important selection criteria in terms of safety, horse-human relationship and performance (e.g. Graf, König von Borstel and Gauly, 2013), the null findings in Chapter 4 suggest personality assessments may not be predictive of the horses' physiological ability to cope with the demands of their situation. This result is coherent with an increasing pool of evidence suggesting that behavioural and physiological reactivity to stressors may not be correlated in the horse (Yarnell, Hall and Billett, 2013; Squibb et al., 2018; Munsters et al., 2013b). This raises concern for individuals who are behaviourally compliant but physiologically stress sensitive, as they may be perceived to be suitable for higher-pressure roles from a behavioural point of view but nevertheless experience repeated activations of the adaptive stress response. This can lead to negative consequences in terms of health (Khansari, Murgo and Faith, 1990) and performance (Bartolomé and Cockram, 2016), thus compromising welfare. This is a concerning finding from a welfare point of view. If confirmed in future studies in which the methodological limitations outlined above are addressed, this finding would greatly limit the relevance of the EPT as an applied selection tool to protect welfare.

In continuity with the suggestion that behavioural indicators of stress do not align with underlying stress physiology, Chapter 5 revealed that more agreeable horses had higher baseline cortisol levels, as measured through hair cortisol concentration. The study design used here did not enable us to establish direct causal links between these two outcomes. Evidence in humans suggests that basal cortisol levels may act as a biological driver of sensitivity to punishment (Honk et al., 2003). Therefore, the findings of this chapter may reflect a potential role of basal cortisol levels as a driver of Agreeableness. Equally, however, the causal link might be reversed. The findings of Chapter 5 suggest that behaviourally compliant individuals do not actually experience reduced physiological responses to stressors. Therefore, it may be hypothesised that increased hair cortisol concentrations in more agreeable horses are instead due to repeated HPA axis activation, as their stress responses are not as easily identified. This hypothesis is supported by recent findings showing that lay equestrians often fail to identify subtle behavioural indicators of stress, especially in contexts in which a stress response is not expected (Bell et al., 2019). Even trained specialists (Equitation scientists and equine veterinarians) differ in their assessment of stress levels based on subtle behavioural indicators and rely preferentially on overt non-compliance to identify a stress response (Pearson et al., 2021). This finding linking Agreeableness with higher chronic HPA axis activity is novel, but in line with the result of previous work showing that personality may confound the expression of negative states in the horse (Ijichi, 2014).

Further research into this line of evidence is warranted due to its applied relevance. Future work should first focus on addressing the limitations of Chapter 4 to confirm whether the EPT has predictive validity for physiological reactivity to stressors. As discussed in Chapter 4, opportunistic data collection during routine husbandry procedures known to be aversive (e.g. trailer loading: Shanahan, 2003; travelling: Fazio et al., 2013) could be used to ethically gather data on physiological responses to highly aversive situations. In addition, the use of gold standard equipment (e.g. portable ECG to record inter-beat intervals) and analytic techniques (e.g. LC-MS for cortisol concentration assay) could help avoid the methodological limitations that affected the results presented in this thesis. In addition, although Chapter 3 has shown Agreeableness to be measured with very high levels of inter-rater and test-retest reliability, Ijichi et al. (2013) could not confirm the predictive validity of this scale. In this study, the behaviours hypothesised to be linked to Agreeableness, such as the time taken to comply with a novel handling task, were instead predicted by Extraversion. Therefore, future work should also focus on establishing whether Agreeableness does reflect the tendency for compliant behaviour in the horse, as this is a crucial element of interpretation of the findings of Chapter 5. To ensure the validity and applied relevance of the construct measured by the Agreeableness factor, it should be validated against behavioural responses to husbandry or equitation procedures.

If future studies confirm the findings of Chapter 4 and Chapter 5, this would raise ethical concerns around selecting horses for roles on the basis of low behavioural reactivity alone. While selection on the basis of behavioural compliance may benefit human safety and the quality of the horse-human relationship, it should be accompanied by increased education to identify subtle signs of stress. This is especially crucial for novice riders or owners who value compliance but likely lack the tools to appropriately assess stress levels in their horses (e.g. Pony Club members: Buckley, Dunn and More, 2004).

7.1.4 PERSONALITY AS A CONFOUNDING FACTOR OF EMERGING MEASURES OF STRESS

Although this was not the primary aim of this work, it should be noted that this thesis has identified personality as a confounding factor of two emerging measures of stress and welfare in the horse. Hair cortisol concentration is emerging as a useful measure of the impact of longer-term challenges such as surgical procedures (Duran et al., 2017) or relocation (Gardela et al., 2020), while spontaneous eye blink rate has been identified as a reliable indicator of subtle, short-term stress responses (Mott, Hawthorne and McBride, 2020; Merkies et al., 2019). This work demonstrates that personality dimensions of the EPT are linked with both hair cortisol (Chapter 5) and spontaneous eye blink rate (Chapter 6). This is in accordance with results from other groups, evidencing links between subjectively assessed personality factors and hair cortisol (Sauveroche et al., 2020) or spontaneous eye blink rate (Roberts et al., 2016). These results are emerging and more work is needed to fully clarify the relationship between personality and hair cortisol concentration or spontaneous blink rate. However, the results of the present work, taken together with the body of work produced by other groups over the duration of this project, strongly suggests that personality can act as a confounding factor of these two emerging measures of stress. Therefore, where these measures are used to monitor or compare welfare across, rather than within, individuals, it is crucial that personality is taken into account in the analysis.

7.2 GENERAL CONCLUSION

The last decades of research in equine personality have given rise to a tentative model of equine personality structure, and have led to the development of objective tests that enable the selection of horses on the basis of personality. Models in other species suggests that differences in personality are driven by underlying neurochemical differences. They therefore co-vary with other traits relevant to equine welfare and performance, such as physiological stress sensitivity and cognitive style. However, little remains known about non-behavioural correlates of equine personality. Thus, it is unclear what impact selection on the basis of personality has on those traits in the horse. In order to start addressing this question, this thesis therefore aimed to explore potential neurophysiological correlates of equine personality. Autonomic and HPA axis reactivity to stressors, chronic HPA axis activity and tonic striatal dopamine levels were chosen as parameters of interest for this preliminary work as they have been identified as correlates of personality in other species, are relevant to equine welfare and performance, and can be investigated fully non-invasively. This thesis resulted in a number of novel findings. Firstly, four out of the five sub-scales of the Equine Personality Test were shown to satisfy all four criteria of validity and reliability expected of a personality assessment tool. No other equine personality questionnaire available in the literature has yet been checked against all four criteria, and this work therefore contributes to position the Equine Personality Test as a highly relevant questionnaire for both research and applied purposes. Secondly, none of the relationships described in other species between personality factors and the physiological parameters of interest could be replicated in the sample of horses used in this work. This may be due in part to limitations in the non-invasive methods used to monitor the physiological parameters of interest; refinements to these methods have been proposed on the basis of the findings of this work. Nonetheless, this finding raises crucial questions regarding the impact of selection on the basis of personality. In particular, the lack of relationship between personality factors and physiological reactivity to stressors suggests that the Equine Personality Assessment does not have predictive validity for physiological stress sensitivity. Alongside this, a positive association between Agreeableness and chronic HPA axis activity was documented. This suggests that more compliant horses may experience more physiological stress responses. From an applied point of view, this suggests that selecting horses on the basis of compliance and limited behavioural reactivity may not be enough to protect their welfare. This is an important finding from an applied point of view. If confirmed by future research, these findings should be used to inform the choice of methods used to select horses for roles on the basis of personality, with a view to maximise not only performance and handler safety but also equine welfare.

Chapter 8 References

Adamczyk, A., Mejias, R., Takamiya, K., Yocum, J., Krasnova, I.N., Calderon, J., Cadet, J.L., Huganir, R.L., Pletnikov, M. V. and Wang, T., 2012. GluA3-deficiency in mice is associated with increased social and aggressive behavior and elevated dopamine in striatum. *Behavioural Brain Research*, 229(1), pp.265–272.

Al-Abdulmunem, M., 1999. Relation between tear breakup time and spontaneous blink rate. *International Contact Lens Clinic*, 26(5), pp.117–120.

Albar, W.F., Russell, E.W., Koren, G., Rieder, M.J. and Van Umm, S.H., 2013. Human hair cortisol analysis: Comparison of the internationally-reported ELISA methods. *Clinical & Investigative Medicine*, [online] 36(6), p.312.

Alboukadel, K., 2020. *'ggplot2' Based Publication Ready Plots*.

Anderson, M.K., Friend, T.H., Evans, J.W. and Bushong, D.M., 1999. Behavioral assessment of horses in therapeutic riding programs. *Applied Animal Behaviour Science*, [online] 63(1), pp.11–24.

Bacher, L.F. and Allen, K.J., 2009. Sensitivity of the rate of spontaneous eye blinking to type of stimuli in young infants. *Developmental Psychobiology*, 51(2), pp.186–197.

Bae, Y.J., Gaudl, A., Jaeger, S., Stadelmann, S., Hiemisch, A., Kiess, W., Willenberg, A., Schaab, M., Von Klitzing, K., Thiery, J., Ceglarek, U., Döhnert, M. and Kratzsch, J., 2016. Immunoassay or LC-MS/MS for the measurement of salivary cortisol in children? *Clinical Chemistry and Laboratory Medicine*, [online] 54(5), pp.811–822.

Banse, H.E., Getachew, F., Levy, M. and Smits, J., 2020. Influence of season and pituitary pars intermedia dysfunction on hair cortisol concentration in horses. *Domestic Animal Endocrinology*, [online] 72, p.106375.

Barbato, G., della Monica, C., Costanzo, A. and De Padova, V., 2012. Dopamine activation in Neuroticism as measured by spontaneous eye blink rate. *Physiology and Behavior*, 105(2), pp.332–336.

Bariselli, S., Hörnberg, H., Prévost-Solié, C., Musardo, S., Hatstatt-Burklé, L., Scheiffele, P. and Bellone, C., 2018. Role of VTA dopamine neurons and neuroligin 3 in sociability traits related to nonfamiliar conspecific interaction. *Nature Communications*, [online] 9(1), pp.1–15.

Bartolomé, E. and Cockram, M.S., 2016. Potential Effects of Stress on the Performance of Sport Horses. *Journal of Equine Veterinary Science*, [online] 40, pp.84–93.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Dai, B., Scheipl, F., Grothendieck, G., Green, P. and Fox, J., 2019. *Linear Mixed-Effects Models using 'Eigen' and S4*.

Bechshøft, T.Ø., Sonne, C., Dietz, R., Born, E.W., Novak, M.A., Henchey, E. and Meyer, J.S., 2011. Cortisol levels in hair of East Greenland polar bears. *Science of The Total Environment*, [online] 409(4), pp.831–834.

Becker-Birck, M., Schmidt, A., Lasarzik, J., Aurich, J., Möstl, E. and Aurich, C., 2013. Cortisol release and heart rate variability in sport horses participating in equestrian competitions. *Journal of Veterinary Behavior: Clinical Applications and Research*, [online] 8(2), pp.87–94.

Bell, C., Rogers, S., Taylor, J. and Busby, D., 2019. Improving the Recognition of Equine Affective States. *Animals*, [online] 9(12), p.1124.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, [online] 57(1), pp.289–300.

Bennett, A. and Hayssen, V., 2010. Measuring cortisol in hair and saliva from dogs: coat color and pigment differences. *Domestic Animal Endocrinology*, [online] 39(3), pp.171–180.

Benus, R.F., Bohus, B., Koolhaas, J.M. and van Oortmerssen, G.A., 1991. Behavioural differences between artificially selected aggressive and non-aggressive mice: response to apomorphine. *Behavioural Brain Research*, 43(2), pp.203–208.

Berenbaum, H. and Williams, M., 1994. Extraversion, hemispatial bias, and eyeblink rates. *Personality and Individual Differences*, 17(6), pp.849–852.

Berenbaum, H. and Williams, M., 1995. Personality and Emotional Reactivity. *Journal of Research in Personality*, 29(1), pp.24–34.

Berghold, P., Möstl, E. and Aurich, C., 2007. Effects of reproductive status and management on cortisol secretion and fertility of oestrous horse mares. *Animal Reproduction Science*, [online] 102(3–4), pp.276–285.

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J. and Kleitman, S., 2019. Detecting personality traits using eye-tracking data. *Conference on Human Factors in Computing Systems - Proceedings*, pp.1–12.

Best, L.J., Alexiades, V., Hendrix, D.V.H., Chen, T. and Ward, D.A., 2018. Blink patterns and kinematics of eyelid motion in ophthalmologically normal horses. *American Journal of Veterinary Research*, 79(6), pp.650–657.

de Boer, S.F., Buwalda, B. and Koolhaas, J.M., 2017. Untangling the neurobiology of coping styles in rodents: Towards neural mechanisms underlying individual differences in disease susceptibility. *Neuroscience & Biobehavioral Reviews*, [online] 74(B), pp.401–422.

De Boer, S.F. and Koolhaas, J.M., 2005. 5-HT1A and 5-HT1B receptor agonists and aggression: A pharmacological challenge of the serotonin deficiency hypothesis. *European Journal of Pharmacology*, 526(1–3), pp.125–139.

Bohák, Z., Harnos, A., Joó, K., Szenci, O. and Kovács, L., 2018. Anticipatory response before competition in Standardbred racehorses. *PLOS ONE*, [online] 13(8), p.e0201691.

Bohák, Z., Szabó, F., Beckers, J.F., Melo de Sousa, N., Kutasi, O., Nagy, K. and Szenci, O., 2013. Monitoring the circadian rhythm of serum and salivary cortisol concentrations in the horse. *Domestic Animal Endocrinology*, 45(1), pp.38–42.

Bohák, Z., Szenci, O., Harnos, A., Kutasi, O. and Kovács, L., 2017. Effect of temperament on cortisol response to a single exercise bout in Thoroughbred racehorses – short communication. *Acta Veterinaria Hungarica*, [online] 65(4), pp.541–545.

Von Borell, E., Langbein, J., Després, G., Hansen, S., Leterrier, C., Marchant-Forde, J., Marchant-Forde, R., Minero, M., Mohr, E., Prunier, A., Valance, D. and Veissier, I., 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*, [online] 92(3), pp.293–316.

Boyle, G.J. and Helmes, E., 2009. Methods of Personality Assessment. In: P.J. Corr and G. Matthews, eds., *The Cambridge Handbook of Personality Psychology*, 1st ed. New York: Cambridge University Press, pp.110–126.

Březinová, V. and Kendell, R.E., 1977. Smooth Pursuit Eye Movements of Schizophrenics and Normal People Under Stress. *British Journal of Psychiatry*, [online] 130(1), pp.59–63.

Briard, L., Dorn, C. and Petit, O., 2015. Personality and Affinities Play a Key Role in the Organisation of Collective Movements in a Group of Domestic Horses. *Ethology*, 121(9), pp.888–902.

Buckley, P., Dunn, T. and More, S.J., 2004. Owners' perceptions of the health and performance of Pony Club horses in Australia. *Preventive Veterinary Medicine*, 63(1–2), pp.121–133.

Bulens, A., Sterken, H., Van Beirendonck, S., Van Thielen, J. and Driessen, B., 2015. The use of different objects during a novel object test in stabled horses. *Journal of Veterinary Behavior: Clinical Applications and Research*, [online] 10(1), pp.54–58.

Burnett, T.A., Madureira, A.M.L., Silper, B.F., Nadalin, A., Tahmasbi, A., Veira, D.M. and Cerri, R.L.A., 2014. Short communication: Factors affecting hair cortisol concentrations in lactating dairy cows. *Journal of Dairy Science*, 97(12), pp.7685–7690.

Cabib, S., 2006. The neurobiology of stereotypy II: the role of stress. In: G. Mason and J. Rushen, eds., *Stereotypic animal behaviour: fundamentals and applications to welfare*. [online] Wallingford: CABI, pp.227–255.

Calcagnoli, F., de Boer, S.F., Beiderbeck, D.I., Althaus, M., Koolhaas, J.M. and Neumann, I.D., 2014. Local oxytocin expression and oxytocin receptor binding in the male rat brain is associated with aggressiveness. *Behavioural Brain Research*, [online] 261, pp.315–322.

Caramaschi, D., de Boer, S.F. and Koolhaas, J.M., 2007. Differential role of the 5-HT1A receptor in aggressive and non-aggressive mice: An across-strain comparison. *Physiology & Behavior*, 90(4), pp.590–601.

Carere, C., Groothuis, T.G.G., Möstl, E., Daan, S. and Koolhaas, J.M., 2003. Fecal corticosteroids in a territorial bird selected for different personalities: daily rhythm and the response to social stress. *Hormones and Behavior*, 43(5), pp.540–548.

Carlitz, E.H.D., Kirschbaum, C., Stalder, T. and van Schaik, C.P., 2014. Hair as a long-term retrospective cortisol calendar in orang-utans (Pongo spp.): New perspectives for stress monitoring in captive management and conservation. *General and Comparative Endocrinology*, [online] 195, pp.151–156.

Carlitz, E.H.D.E.H.D., Kirschbaum, C., Miller, R., Rukundo, J. and van Schaik, C.P.C.P., 2015. Effects of body region and time on hair cortisol concentrations in chimpanzees (Pan troglodytes). *General and Comparative Endocrinology*, [online] 223, pp.9–15.

Carnevali, L., Nalivaiko, E. and Sgoifo, A., 2014. Respiratory patterns reflect different levels of aggressiveness and emotionality in Wild-type Groningen rats. *Respiratory Physiology & Neurobiology*, 204, pp.28–35.

Carnevali, L., Trombini, M., Porta, A., Montano, N., Boer, S.F. de and Sgoifo, A., 2013. Vagal Withdrawal and Susceptibility to Cardiac Arrhythmias in Rats with High Trait Aggressiveness. *PLOS ONE*, [online] 8(7), p.e68316.

Cavanagh, J.F., Masters, S.E., Bath, K. and Frank, M.J., 2014. Conflict acts as an implicit cost in reinforcement learning. *Nature Communications*, [online] 5(1), pp.1–10.

Chen, E.Y.H., Lam, L.C.W., Chen, R.Y.L. and Nguyen, D.G.H., 1996. Blink rate, neurocognitive impairments, and symptoms in schizophrenia. *Biological Psychiatry*, [online] 40(7), pp.597–603.

Cherry, R.L., Adair, H.S., Chen, T., Hendrix, D.V.H. and Ward, D.A., 2020. Effect of attentional focus levels on spontaneous eyeblink rate in horses. *Veterinary Ophthalmology*, [online] 23(4), pp.690–695.

Chida, Y. and Hamer, M., 2008. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: A quantitative review of 30 years of investigations. *Psychological Bulletin*, [online] 134(6), pp.829–885.

Chida, Y. and Steptoe, A., 2009. Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. *Biological Psychology*, 80(3), pp.265–278.

Christensen, J.W., Keeling, L.J. and Nielsen, B.L., 2005. Responses of horses to novel visual, olfactory and auditory stimuli. *Applied Animal Behaviour Science*, [online] 93(1–2), pp.53–65.

Christensen, J.W., Malmkvist, J., Nielsen, B.L. and Keeling, L.J., 2008. Effects of a calm companion on fear reactions in naive test horses. *Equine Veterinary Journal*, [online] 40(1), pp.46–50.

Christensen, J.W., Zharkikh, T. and Chovaux, E., 2011. Object recognition and generalisation during habituation in horses. *Applied Animal Behaviour Science*, [online] 129(2–4), pp.83–91.

Christensen, J.W., Zharkikh, T. and Ladewig, J., 2008. Do horses generalise between objects during habituation? *Applied Animal Behaviour Science*, [online] 114(3–4), pp.509–520.

Christensen, J.W.W., Beekmans, M., van Dalum, M. and VanDierendonck, M., 2014. Effects of hyperflexion on acute stress responses in ridden dressage horses. *Physiology and Behavior*, [online] 128, pp.39–45.

Cloninger, C.R., Svrakic, D.M. and Przybeck, T.R., 1993. A Psychobiological Model of Temperament and Character. *Archives of General Psychiatry*, [online] 50(12), pp.975–990.

Colzato, L.S., Slagter, H.A., van den Wildenberg, W.P.M. and Hommel, B., 2009. Closing one's eyes to reality: Evidence for a dopaminergic basis of Psychoticism from spontaneous eye blink rates. *Personality and Individual Differences*, 46(3), pp.377–380.

Comin, A., Veronesi, M.C., Montillo, M., Faustini, M., Valentini, S., Cairoli, F. and Prandi, A., 2012. Hair cortisol level as a retrospective marker of hypothalamic-pituitary-adrenal axis activity in horse foals. *Veterinary Journal*, 194(1), pp.131–132.

Cooke, R.F., Schubach, K.M., Marques, R.S., Peres, R.F.G., Silva, L.G.T., Carvalho, R.S., Cipriano, R.S., Bohnert, D.W., Pires, A. V. and Vasconcelos, J.L.M., 2017. Effects of temperament on physiological, productive, and reproductive responses in Bos indicus beef cows. *Journal of Animal Science*, [online] 95(1), pp.1–8.

Coppens, C.M., de Boer, S.F. and Koolhaas, J.M., 2010. Coping styles and behavioural flexibility: towards underlying mechanisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, [online] 365(1560), pp.4021–4028.

Cori, J.M., Anderson, C., Shekari Soleimanloo, S., Jackson, M.L. and Howard, M.E., 2019. Narrative review: Do spontaneous eye blink parameters provide a useful assessment of state drowsiness? *Sleep Medicine Reviews*, [online] 45, pp.95–104.

Cortina, J.M., 1993. What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), pp.98–104.

Costa, P.T.J. and Mccrae, R.R., 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Ressources.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), pp.297–334.

Dang, L.C., Samanez-Larkin, G.R., Castrellon, J.J., Perkins, S.F., Cowan, R.L., Newhouse, P.A. and Zald, D.H., 2017. Spontaneous Eye Blink Rate (EBR) Is Uncorrelated with Dopamine D2 Receptor Availability and Unmodulated by Dopamine Agonism in Healthy Adults. *eneuro*, [online] 4(5).

Davenport, M.D.M.D., Tiefenbacher, S., Lutz, C.K.C.K., Novak, M.A.M.A. and Meyer, J.S.J.S., 2006. Analysis of endogenous cortisol concentrations in the hair of rhesus macaques. *General and Comparative Endocrinology*, [online] 147(3), pp.255–261.

Depue, R.A. and Collins, P.F., 1999. Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, [online] 22(3), pp.491–517.

Depue, R.A. and Morrone-Strupinsky, J. V., 2005. A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, [online] 28(3), pp.313–395.

Deyoung, C.G., 2006. Higher-order factors of the big five in a multi-informant sample. *Journal* *of Personality and Social Psychology*, 91(6), pp.1138–1151.

DeYoung, C.G., 2013. The neuromodulator of exploration: A unifying theory of the role of dopamine in personality. *Frontiers in Human Neuroscience*, [online] 7(NOV), p.762.

DeYoung, C.G. and Gray, J.R., 2009. Personality neuroscience: explaining individual differences in affect, behaviour and cognition. In: P.J. Corr and G. Matthews, eds., *The Cambridge Handbook of Personality Psychology*, 1st ed. New York: Cambridge University Press, pp.323–346.

DeYoung, C.G., Hirsh, J.B., Shane, M.S., Papademetris, X., Rajeevan, N. and Gray, J.R., 2010. Testing Predictions From Personality Neuroscience. *Psychological Science*, [online] 21(6), pp.820–828.

Dingemanse, N.J. and Wright, J., 2020. Criteria for acceptable studies of animal personality and behavioural syndromes. *Ethology*, [online] 126(9), pp.865–869.

Doughty, M.J., 2001. Consideration of three types of spontaneous eyeblink activity in normal humans: During reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science*, [online] 78(10), pp.712–725.

Doughty, M.J., 2016. Assessment of short-term variability in human spontaneous blink rate during video observation with or without head / chin support. *Clinical and Experimental Optometry*, [online] 99(2), pp.135–141.

Dulude-de Broin, F., Côté, S.D., Whiteside, D.P. and Mastromonaco, G.F., 2019. Faecal metabolites and hair cortisol as biological markers of HPA-axis activity in the Rocky mountain goat. *General and Comparative Endocrinology*, 280, pp.147–157.

Duran, M.C., Janz, D.M., Waldner, C.L., Campbell, J.R. and Marques, F.J., 2017. Hair Cortisol Concentration as a Stress Biomarker in Horses: Associations With Body Location and Surgical Castration. *Journal of Equine Veterinary Science*, [online] 55, pp.27–33.

Ellis, A.D., Stephenson, M., Preece, M. and Harris, P., 2014. A novel approach to systematically compare behavioural patterns between and within groups of horses. *Applied Animal Behaviour Science*, [online] 161, pp.60–74.

Elsworth, J.D., Lawrence, M.S., Roth, R.H., Taylor, J.R., Mailman, R.B., Nichols, D.E., Lewis, M.H. and Redmond, D.E., 1991. D1 and D2 Dopamine Receptors Independently Regulate Spontaneous Blink Rate in the Vervet Monkey. *Journal of Pharmacology and Experimental Therapeutics*, 259(2), pp.595–600.

Eser, H.P., Pötsch, L., Skopp, G. and Moeller, M.R., 1997. Influence of sample preparation on analytical results: Drug analysis [GC/MS] on hair snippets versus hair powder using various extraction methods. *Forensic Science International*, 84(1–3), pp.271–279.

Everts, H.G.J. and Koolhaas, J.M., 1999. Differential modulation of lateral septal vasopressin receptor blockade in spatial learning, social recognition, and anxiety-related behaviors in rats. *Behavioural Brain Research*, 99(1), pp.7–16.

Ewacha, M.V.A., Roth, J.D., Anderson, W.G., Brannen, D.C. and Dupont, D.L.J., 2017. Disturbance and chronic levels of cortisol in boreal woodland caribou. *The Journal of Wildlife Management*, [online] 81(7), pp.1266–1275.

Eysenck, H.J., 1983. Psychophysiology and Personality: Extraversion, Neuroticism and Psychoticism. *Individual Differences and Psychopathology*, 3, pp.13–30.

Eysenck, H.J. and Eysenck, S.B.G., 1975. *Manual of the Eysenck Personality Questionnaire (Junior and Adult)*. Kent, UK: Hodder & Stoughton.

Eysenck, S.B.G., Eysenck, H.J. and Barrett, P., 1985. A revised version of the psychoticism scale. *Personality and Individual Differences*, [online] 6(1), pp.21–29.

Fazio, E., Medica, P., Cravana, C. and Ferlazzo, A., 2013. Cortisol response to road transport stress in calm and nervous stallions. *Journal of Veterinary Behavior: Clinical Applications and*

Research, [online] 8(4), pp.231–237.

Ferrari, C., Pasquaretta, C., Carere, C., Cavallone, E., von Hardenberg, A. and Réale, D., 2013. Testing for the presence of coping styles in a wild mammal. *Animal Behaviour*, [online] 85(6), pp.1385–1396.

Field, A., 2009. *Discovering statistics using SPSS*. 3rd ed. *Discovering statistics using SPSS*. London: Sage.

Finkemeier, M.-A., Langbein, J. and Puppe, B., 2018. Personality Research in Mammalian Farm Animals: Concepts, Measures, and Relationship to Welfare. *Frontiers in Veterinary Science*, [online] 5, p.131.

Fitzpatrick, E., Hohl, N., Silburn, P., O'Gorman, C. and Broadley, S.A., 2012. Case-control study of blink rate in Parkinson's disease under different conditions. *Journal of Neurology*, [online] 259(4), pp.739–744.

Fokkema, D.S., Koolhaas, J.M. and van der Gugten, J., 1995. Individual characteristics of behavior, blood pressure, and adrenal hormones in colony rats. *Physiology & Behavior*, 57(5), pp.857–862.

Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E. and Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. *Physiology & Behavior*, [online] 92(3), pp.340–374.

Fourie, N.H., Brown, J.L., Jolly, C.J., Phillips-Conroy, J.E., Rogers, J. and Bernstein, R.M., 2016. Sources of variation in hair cortisol in wild and captive non-human primates. *Zoology*, [online] 119, pp.119–125.

Fourie, N.H., Jolly, C.J., Phillips-Conroy, J.E., Brown, J.L. and Bernstein, R.M., 2015. Variation of hair cortisol concentrations among wild populations of two baboon species (Papio anubis, P. hamadryas) and a population of their natural hybrids. *Primates*, [online] 56(3), pp.259–

272.

French, J.M., 1993. Assessment of donkey temperament and the influence of home environment. *Applied Animal Behaviour Science*, 36(2–3), pp.249–257.

Friard, O. and Gamba, M., 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, [online] 7(11), pp.1325–1330.

Funder, D.C., 1995. On the Accuracy of Personality Judgment: A Realistic Approach. *Psychological Review*, 102(4), pp.652–670.

Funder, D.C., Kolar, D.C. and Blackman, M.C., 1995. Agreement Among Judges of Personality: Interpersonal Relations, Similarity, and Acquaintanceship. *Journal of Personality and Social Psychology*, 69(4), pp.656–672.

Gamer, M., Lemon, J. and Fellows Puspendra Singh, I., 2019. *irr: Various Coefficients of Interrater Reliability and Agreement*.

Garcia, D., Lester, N., Cloninger, K.M. and Robert Cloninger, C., 2017. Temperament and Character Inventory (TCI). In: V. Zeigler-Hill and T.K. Shackelford, eds., *Encyclopedia of Personality and Individual Differences*. [online] Cham: Springer International Publishing, pp.1–3.

Gardela, J., Carbajal, A., Tallo-Parra, O., Olvera-Maneu, S., Álvarez-Rodríguez, M., Jose-Cunilleras, E. and López-Béjar, M., 2020. Temporary Relocation during Rest Periods: Relocation Stress and Other Factors Influence Hair Cortisol Concentrations in Horses. *Animals*, [online] 10(4), p.642.

González-de-la-Vara, M. del R., Valdez, R.A., Lemus-Ramirez, V., Vázquez-Chagoyán, J.C., Villa-Godoy, A. and Romano, M.C., 2011. Effects of adrenocorticotropic hormone challenge and age on hair cortisol concentrations in dairy cattle. *Canadian journal of veterinary* *research = Revue canadienne de recherche veterinaire*, [online] 75(3), pp.216–21.

Górecka-Bruzda, A., Chruszczewski, M.H., Jaworski, Z., Golonka, M., Jezierski, T., Dlugosz, B. and Pieszka, M., 2011. Looking for an Ideal Horse: Rider preferences. *Anthrozoos*, 24(4), pp.379–392.

Gosling, S.D., 2001. From mice to men - What can we learn about personality from animal research. *Psychological Bulletin*, 127(1), pp.45–86.

Gosling, S.D., 2008. Personality in Non-human Animals. *Social and Personality Psychology Compass*, 22(10), pp.985–1001.

Gosling, S.D. and John, O.P., 1999. Personality Dimensions in Nonhuman Animals. *Current Directions in Psychological Science*, [online] 8(3), pp.69–75.

Gosling, S.D., Kwan, V.S.Y. and John, O.P., 2003. A Dog's Got Personality: A Cross-Species Comparative Approach to Personality Judgments in Dogs and Humans. *Journal of Personality and Social Psychology*, 85(6), pp.1161–1169.

Gosling, S.D. and Vazire, S., 2002. Are we barking up the right tree? Evaluating a comparative approach to personality. *Journal of Research in Personality*, 36(6), pp.607–614.

Gow, R., Thomson, S., Rieder, M., Van Uum, S. and Koren, G., 2010. An assessment of cortisol analysis in hair and its clinical applications. *Forensic Science International*, [online] 196(1–3), pp.32–37.

Graf, P., König von Borstel, U. and Gauly, M., 2013. Importance of personality traits in horses to breeders and riders. *Journal of Veterinary Behavior*, 8(5), pp.316–325.

Graf, P., König von Borstel, U. and Gauly, M., 2014. Practical considerations regarding the implementation of a temperament test into horse performance tests: Results of a large-scale test run. *Journal of Veterinary Behavior: Clinical Applications and Research*, [online] 9(6), pp.329–340.

Graf, P., Schneider, T., König von Borstel, U. and Gauly, M., 2013. Economic evaluation of an objective temperament assessment in horses. *Züchtungskunde*, [online] 85(2), pp.129–142.

Grandin, T., Fine, A.H. and Bowers, C., 2010. The use of therapy animals with individuals with autism spectrum disorders. *Handbook on Animal-Assisted Therapy*, pp.247–264.

Grant, K.S.K.S., Worlein, J.M.J.M., Meyer, J.S.J.S., Novak, M.A.M.A., Kroeker, R., Rosenberg, K., Kenney, C. and Burbacher, T.M.T.M., 2017. A longitudinal study of hair cortisol concentrations in *Macaca nemestrina* mothers and infants. *American Journal of Primatology*, [online] 79(2), p.e22591.

Groothuis, T.G.G. and Carere, C., 2005. Avian personalities: characterization and epigenesis. *Neuroscience & Biobehavioral Reviews*, 29(1), pp.137–150.

Gros, D.F., 2011. Startle Inhibition to Positive-Activated Compared to Neutral Stimuli: Variations in Self-Reported Behavioral Approach. *Journal of Psychopathology and Behavioral Assessment*, [online] 33(3), pp.308–314.

Hada, T., Onaka, T., Kusunose, R. and Yagi, K., 2001. Effects of Novel Environmental Stimuli on Neuroendocrine Activity in Thoroughbred Horses. *Journal of Equine Science*, [online] 12(2), pp.33–38.

Hall, C., Goodwin, D., Heleski, C., Randle, H. and Waran, N., 2008. Is There Evidence of Learned Helplessness in Horses? *Journal of Applied Animal Welfare Science*, [online] 11(3), pp.249–266.

Hall, C., Kay, R. and Yarnell, K., 2014. Assessing ridden horse behavior: Professional judgment and physiological measures. *Journal of Veterinary Behavior*, [online] 9(1), pp.22–29.

Hallgren, K.A., 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), pp.23–34.

Hamel, A.F., Lutz, C.K., Coleman, K., Worlein, J.M., Peterson, E.J., Rosenberg, K.L., Novak, M.A. and Meyer, J.S., 2017. Responses to the Human Intruder Test are related to hair cortisol phenotype and sex in rhesus macaques (Macaca mulatta). *American Journal of Primatology*, [online] 79(1), p.e22526.

Hamel, A.F., Meyer, J.S., Henchey, E., Dettmer, A.M., Suomi, S.J. and Novak, M.A., 2011. Effects of shampoo and water washing on hair cortisol concentrations. *Clinica Chimica Acta*, [online] 412(3–4), pp.382–385.

Hausberger, M., Bruderer, C., Le Scolan, N., Pierre, J.-S. and De Rennes, U., 2004. Interplay Between Environmental and Genetic Factors in Temperament/ Personality Traits in Horses (Equus caballus). *Journal of Comparative Psychology*, [online] 118(4), pp.434–446.

Hausberger, M., Muller, C. and Lunel, C., 2011. Does Work Affect Personality? A Study in Horses. *PLoS ONE*, [online] 6(2), p.e14659.

Heimbürge, S., Kanitz, E. and Otten, W., 2019. *The use of hair cortisol for the assessment of stress in animals*. *General and Comparative Endocrinology*, .

Heimbürge, S., Kanitz, E., Tuchscherer, A. and Otten, W., 2020. Within a hair’s breadth – Factors influencing hair cortisol levels in pigs and cattle. *General and Comparative Endocrinology*, 288, p.113359.

Hemmings, A., McBride, S.D. and Hale, C.E., 2007. Perseverative responding and the aetiology of equine oral stereotypy. *Applied Animal Behaviour Science*, [online] 104, pp.143–150.

Henderson, G.L., 1993. Mechanisms of drug incorporation into hair. *Forensic Science International*, 63(1–3), pp.19–29.

Henry, J.P. and Stephens, P.M., 1977. *Stress, Health, and the Social Environment : a Sociobiologic Approach to Medicine*. Springer New York.

Hernández-Vicente, A., Hernando, D., Marín-Puyalto, J., Vicente-Rodríguez, G., Garatachea, N., Pueyo, E. and Bailón, R., 2021. Validity of the polar h7 heart rate sensor for heart rate variability analysis during exercise in different age, body composition and fitness level groups. *Sensors (Switzerland)*, [online] 21(3), pp.1–14.

Hessing, M.J.., Hagelsø, A.M., Schouten, W.G.., Wiepkema, P.R. and Van Beek, J.A.., 1994. Individual behavioral and physiological strategies in pigs. *Physiology & Behavior*, [online] 55(1), pp.39–46.

Hockenhull, J. and Creighton, E., 2013. The use of equipment and training practices and the prevalence of owner-reported ridden behaviour problems in UK leisure horses. *Equine Veterinary Journal*, 45(1), pp.15–19.

Honk, J. van, Schutter, D.J.L.G., Hermans, E.J. and Putman, P., 2003. Low cortisol levels and the balance between punishment sensitivity and reward dependency. *NeuroReport*, [online] 14(15), pp.1993–1996.

Hori, Y., Tozaki, T., Nambo, Y., Sato, F., Ishimaru, M., Inoue-Murayama, M. and Fujita, K., 2016. Evidence for the effect of serotonin receptor 1A gene (HTR1A) polymorphism on tractability in Thoroughbred horses. *Animal Genetics*, 47(1), pp.62–67.

Hotchkiss, J.W., Reid, S.W.J. and Christley, R.M., 2007. A survey of horse owners in Great Britain regarding horses in their care. Part 1: Horse demographic characteristics and management. *Equine Veterinary Journal*, 39(4), pp.294–300.

Hutson, G.D. and Haskell, M.J., 1997. Pre-race behaviour of horses as a predictor of race finishing order. *Applied Animal Behaviour Science*, 53(4), pp.231–248.

Ijichi, C., Collins, L.M., Creighton, E. and Elwood, R.W., 2013. Harnessing the power of personality assessment: subjective assessment predicts behaviour in horses. *Behavioural Processes*, [online] 96, pp.47–52.

Ijichi, C., Collins, L.M. and Elwood, R.W., 2014. Pain expression is linked to personality in horses. *Applied Animal Behaviour Science*, [online] 152, pp.38–43.

Ijichi, C., Wild, H., Dai, F., Bordin, A., Cameron-Whytock, H., White, S.J., Yarnell, K., Starbuck, G., Jolivald, A., Birkbeck, L., Hallam, S. and Costa, E.D., 2020. Dually investigated: The effect of a pressure headcollar on the behaviour, discomfort and stress of trained horses. *Applied Animal Behaviour Science*, 232, p.105101.

Ijichi, C.L., 2014. *Does personality affect the expression of negative state in horses?* Queen's University Belfast.

Ille, N., Aurich, J., Erber, R. and Aurich, C., 2014. Comparison of heart rate and heart rate variability obtained by heart rate monitors and simultaneously recorded electrocardiogram signals in nonexercising horses. *Journal of Veterinary Behavior: Clinical Applications and Research*, 9(6), pp.341–346.

Inoue-Murayama, M., Yokoyama, C., Yamanashi, Y. and Weiss, A., 2018. Common marmoset (Callithrix jacchus) personality, subjective well-being, hair cortisol level and AVPR1a, OPRM1, and DAT genotypes. *Scientific Reports*, [online] 8(1), p.10255.

Iwanicki, S. and Lehmann, J., 2015. Behavioral and trait rating assessments of personality in common marmosets (Callithrix jacchus). *Journal of Comparative Psychology*, [online] 129(3), pp.205–217.

Jacobson, S.L., Freeman, H.D., Santymire, R.M. and Ross, S.R., 2017. Atypical experiences of captive chimpanzees (Pan troglodytes) are associated with higher hair cortisol concentrations as adults. *Royal Society Open Science*, 4(12).

John, O.P. and Robins, R.W., 1993. Determinants of Interjudge Agreement on Personality

Traits: The Big Five Domains, Observability, Evaluativeness, and the Unique Perspective of the Self. *Journal of Personality*, 61(4), pp.521–551.

Jongkees, B.J. and Colzato, L.S., 2016. Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews*, [online] 71, pp.58–82.

Kaminer, J., Powers, A.S., Horn, K.G., Hui, C. and Evinger, C., 2011. Characterizing the spontaneous blink generator: An animal model. *Journal of Neuroscience*, [online] 31(31), pp.11256–11267.

Karson, C.N., 1983. Spontaneous eye-blink rate and dopaminergic systems. *Brain*, [online] 106(3), pp.643–653.

Khansari, D.N., Murgo, A.J. and Faith, R.E., 1990. Effects of stress on the immune system. *Immunology Today*, 11(C), pp.170–175.

Khoury, J.E., Gonzalez, A., Levitan, R.D., Pruessner, J.C., Chopra, K., Basile, V.S., Masellis, M., Goodwill, A. and Atkinson, L., 2015. Summary cortisol reactivity indicators: Interrelations and meaning. *Neurobiology of Stress*, 2, pp.34–43.

King, J.E. and Figueredo, A.J., 1997. The Five-Factor Model plus Dominance in Chimpanzee Personality. *Journal of Research in Personality*, 31(2), pp.257–271.

Kirschbaum, C., Pirke, K.-M. and Hellhammer, D.H., 1993. The ‘Trier Social Stress Test’ – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, [online] 28(1–2), pp.76–81.

König von Borstel, U., 2013. Assessing and influencing personality for improvement of animal welfare: a review of equine studies. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 8(6), pp.7–21.

König von Borstel, U., Euent, S., Graf, P., König, S. and Gauly, M., 2011. Equine behaviour and heart rate in temperament tests with or without rider or handler. *Physiology & Behavior,* [online] 104(3), pp.454–463.

König von Borstel, U., Pasing, S., Gauly, M. and Christmann, L., 2013. Status quo of the personality trait evaluation in horse breeding: Judges' assessment of the situation and strategies for improvement. *Journal of Veterinary Behavior: Clinical Applications and Research*, [online] 8, pp.326–334.

Koo, T.K. and Li, M.Y., 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, [online] 15(2), pp.155–163.

Koolhaas, J.., Korte, S.., De Boer, S.., Van Der Vegt, B.., Van Reenen, C.., Hopster, H., De Jong, I.., Ruis, M.A.. and Blokhuis, H.., 1999. Coping styles in animals: current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, [online] 23(7), pp.925–935.

Koolhaas, J.M., 2008. Coping style and immunity in animals: Making sense of individual variation. *Brain, Behavior, and Immunity*, [online] 22(5), pp.662–667.

Koolhaas, J.M., De Boer, S.F., Coppens, C.M. and Buwalda, B., 2010. Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology*, [online] 31(3), pp.307–321.

Koren, L., Mokady, O., Karaskov, T., Klein, J., Koren, G. and Geffen, E., 2002. A novel method using hair for determining hormonal levels in wildlife. *Animal Behaviour*, 63(2), pp.403–406.

Korte, S.M., Beuving, G., Ruesink, W. and Blokhuis, H.J., 1997. Plasma Catecholamine and Corticosterone Levels During Manual Restraint in Chicks from a High and Low Feather Pecking Line of Laying Hens. *Physiology & Behavior*, 62(3), pp.437–441.

Korte, S.M., Bouws, G.A.H., Koolhaas, J.M. and Bohus, B., 1992. Neuroendocrine and behavioral responses during conditioned active and passive behavior in the defensive burying/probe avoidance paradigm: Effects of ipsapirone. *Physiology & Behavior*, 52(2), pp.355–361.

Korte, S.M., Meijer, O.C., De Kloet, E.R., Buwalda, B., Keijser, J., Sluyter, F., Van Oortmerssen, G. and Bohus, B., 1996. Enhanced 5-HT1A receptor expression in forebrain regions of aggressive house mice. *Brain Research*, 736(1–2), pp.338–343.

Korte, S.M., Ruesink, W. and Blokhuis, H.J., 1998. Heart Rate Variability During Manual Restraint in Chicks From High- and Low-Feather Pecking Lines of Laying Hens. *Physiology & Behavior*, [online] 65(4–5), pp.649–652.

Koski, S.E., Buchanan-Smith, H.M., Ash, H., Burkart, J.M., Bugnyar, T. and Weiss, A., 2017. Common marmoset (Callithrix jacchus) personality. *Journal of Comparative Psychology*, [online] 131(4), pp.326–336.

Kralj-Fišer, S. and Schuett, W., 2014. Studying personality variation in invertebrates: why bother? *Animal Behaviour*, 91, pp.41–52.

Kroshko, T., Kapronczai, L., Cattet, M.R.L.M.R.L., Macbeth, B.J.B.J., Stenhouse, G.B.G.B., Obbard, M.E.M.E. and Janz, D.M.D.M., 2017. Comparison of methanol and isopropanol as wash solvents for determination of hair cortisol concentration in grizzly bears and polar bears. *MethodsX*, [online] 4, pp.68–75.

Lafferty, D.J.R., Laudenslager, M.L., Mowat, G., Heard, D. and Belant, J.L., 2015. Sex, Diet, and the Social Environment: Factors Influencing Hair Cortisol Concentration in Free-Ranging Black Bears (Ursus americanus). *PLOS ONE*, [online] 10(11), p.e0141489.

Lansade, L. and Bouissou, M.-F., 2008. Reactivity to humans: A temperament trait of horses which is stable across time and situations. *Applied Animal Behaviour Science*, [online] 114(3–4), pp.492–508.

Lansade, L., Bouissou, M.-F. and Erhard, H.W., 2007. Reactivity to isolation and association with conspecifics: A temperament trait stable across time and situations. *Applied Animal Behaviour Science*, [online] 109(2–4), pp.355–373.

Lansade, L., Bouissou, M.-F. and Erhard, H.W., 2008. Fearfulness in horses: A temperament trait stable across time and situations. *Applied Animal Behaviour Science*, [online] 115(3–4), pp.182–200.

Lansade, L., Marchand, A.R., Coutureau, E., Ballé, C., Polli, F. and Calandreau, L., 2017. Personality and predisposition to form habit behaviours during instrumental conditioning in horses (Equus caballus). *PLoS ONE*, [online] 12(2), p.e0171010.

Lansade, L., Philippon, P., Hervé, L. and Vidament, M., 2016. Development of personality tests to use in the field, stable over time and across situations, and linked to horses' show jumping performance. *Applied Animal Behaviour Science*, [online] 176, pp.43–51.

Lansade, L., Pichard, G. and Leconte, M., 2008. Sensory sensitivities: Components of a horse's temperament dimension. *Applied Animal Behaviour Science*, [online] 114(3–4), pp.534–553.

Laudenslager, M.L.M.L., Jorgensen, M.J.M.J., Grzywa, R. and Fairbanks, L.A.L.A., 2011. A novelty seeking phenotype is related to chronic hypothalamic-pituitary-adrenal activity reflected by hair cortisol. [online] 104(2), pp.291–295.

Lawrence, M.S. and Redmond, D.E., 1991. MPTP lesions and dopaminergic drugs alter eye blink rate in African green monkeys. *Pharmacology Biochemistry and Behavior*, [online] 38(4), pp.869–874.

Lean, Y. and Shan, F., 2012. Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, [online] 22(3), pp.177–187.

Leiner, L. and Fendt, M., 2011. Behavioural fear and heart rate responses of horses after exposure to novel objects: Effects of habituation. *Applied Animal Behaviour Science*, [online] 131(3–4), pp.104–109.

Lenoir, A., Trachsel, D.S., Younes, M., Barrey, E. and 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*, [online] 4(OCT), p.17.

Lesimple, C., Fureix, C., LeScolan, N., Richard-Yris, M.A. and Hausberger, M., 2011. Housing conditions and breed are associated with emotionality and cognitive abilities in riding school horses. *Applied Animal Behaviour Science*, 129(2–4), pp.92–99.

Ley, J.M., Bennett, P.C. and Coleman, G.J., 2009. A refinement and validation of the Monash Canine Personality Questionnaire (MCPQ). *Applied Animal Behaviour Science*, 116(2–4), pp.220–227.

Ley, J.M., McGreevy, P. and Bennett, P.C., 2009. Inter-rater and test-retest reliability of the Monash Canine Personality Questionnaire-Revised (MCPQ-R). *Applied Animal Behaviour Science*, 119(1–2), pp.85–90.

Lloyd, A.S., Martin, J.E., Bornett-Gauci, H.L.I. and Wilkinson, R.G., 2008. Horse personality: Variation between breeds. *Applied Animal Behaviour Science*, 112, pp.369–383.

Lloyd, A.S., Martin, J.E., Louise, H., Bornett-Gauci, I. and Wilkinson, R.G., 2007. Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. *Applied Animal Behaviour Science*, [online] 105(1–3), pp.205–222.

Loasby, S., 2018. *An investigation into the relationship between personality factors and cognitive abilities in the domestic horse*. Nottingham Trent University.

Lockwood, S.A., Kattesh, H.G., Rhinehart, J.D., Strickland, L.G., Krawczel, P.D., Wilkerson, J.B., Kirkpatrick, F.D. and Saxton, A.M., 2017. Relationships among temperament, acute and chronic cortisol and testosterone concentrations, and breeding soundness during performance testing of Angus bulls. *Theriogenology*, [online] 89, pp.140–145.

Macbeth, B.J., Cattet, M.R.L., Stenhouse, G.B., Gibeau, M.L. and Janz, D.M., 2010. Hair cortisol concentration as a noninvasive measure of long-term stress in free-ranging grizzly bears (Ursus arctos): Considerations with implications for other wildlife. *Canadian Journal of Zoology*, 88(10), pp.935–949.

Maia, T. V. and Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, [online] 14(2), pp.154–162.

Malcolm, K.D., McShea, W.J., Van Deelen, T.R., Bacon, H.J., Liu, F., Putman, S., Zhu, X. and Brown, J.L., 2013. Analyses of fecal and hair glucocorticoids to evaluate short- and long-term stress and recovery of Asiatic black bears (Ursus thibetanus) removed from bile farms in China. *General and Comparative Endocrinology*, [online] 185, pp.97–106.

Martin, J.G.A. and Réale, D., 2008. Animal temperament and human disturbance: Implications for the response of wildlife to tourism. *Behavioural Processes*, [online] 77(1), pp.66–72.

McBride, S.D. and Hemmings, A., 2005. Altered mesoaccumbens and nigro-striatal dopamine physiology is associated with stereotypy development in a non-rodent species. *Behavioural Brain Research*, [online] 159, pp.113–118.

McBride, S.D. and Mills, D.S., 2012. Psychological factors affecting equine performance. *BMC Veterinary Research*, [online] 8, p.180.

McBride, S.D., Parker, M.O., Roberts, K. and Hemmings, A., 2017. Applied neurophysiology of the horse; implications for training, husbandry and welfare. *Applied Animal Behaviour Science*, [online] 190, pp.90–101.

Mccall, C.A.A., Hall, S., McElhenney, W.H.H. and Cummins, K.A.A., 2006. Evaluation and comparison of four methods of ranking horses based on reactivity. *Applied Animal Behaviour Science*, [online] 96(1–2), pp.115–127.

Mccrae, R.R. and Costa, P.T., 1987. Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*, [online] 52(1), pp.81–90.

Mccrae, R.R. and Costa, P.T., 2004. A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, [online] 36(3), pp.587–596.

McCrae, R.R., Costa, P.T., Hřebíčkovă, M., Ostendorf, F., Angleitner, A., Avia, M.D., Sanz, J., Sánchez-Bernardos, M.L., Kusdil, M.E., Woodfield, R., Saunders, P.R. and Smith, P.B., 2000. Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, 78(1), pp.173–186.

McCrae, R.R., Costa, P.T., Martin, T.A., Oryol, V.E., Rukavishnikov, A.A., Senin, I.G., Hřebíčková, M. and Urbánek, T., 2004. Consensual validation of personality traits across cultures. *Journal of Research in Personality*, 38(2), pp.179–201.

McDonnell, S., 2003. *Practical field guide to horse behavior: the equid ethogram. Practical field guide to horse behavior: the equid ethogram.* Lexington, USA: The Blood-Horse, Inc.

McGraw, K.O. and Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, [online] 1(1), pp.30–46.

McGreevy, P.D. and McLean, A.N., 2007. Roles of learning theory and ethology in equitation. *Journal of Veterinary Behavior*, 2(4), pp.108–118.

McGrogan, C., Hutchison, M.D. and King, J.E., 2008. Dimensions of horse personality based on owner and trainer supplied personality traits. *Applied Animal Behaviour Science*, 13(1–3), pp.206–214.

McLean, A.N. and McGreevy, P.D., 2010. Horse-training techniques that may defy the principles of learning theory and compromise welfare. *Journal of Veterinary Behavior*, 5(4), pp.187–195.

Mehta, P.H., Welker, K.M., Zilioli, S. and Carré, J.M., 2015. Testosterone and cortisol jointly modulate risk-taking. *Psychoneuroendocrinology*, [online] 56, pp.88–99.

Merkies, K., Ready, C., Farkas, L. and Hodder, A., 2019. Eye Blink Rates and Eyelid Twitches as a Non-Invasive Measure of Stress in the Domestic Horse. *Animals*, [online] 9(8), p.562.

Merl, S., Scherzer, S., Palme, R. and Möstl, E., 2000. Pain causes increased concentrations of glucocorticoid metabolites in horse feces. *Journal of Equine Veterinary Science*, 20(9), pp.586–590.

Meyer, J., Novak, M., Hamel, A. and Rosenberg, K., 2014. Extraction and analysis of cortisol from human and monkey hair. *Journal of visualized experiments : JoVE*, [online] (83), p.e50882.

Meyer, J.S. and Novak, M.A., 2012. Minireview: Hair Cortisol: A Novel Biomarker of Hypothalamic-Pituitary-Adrenocortical Activity. *Endocrinology*, [online] 153(9), pp.4120–4127.

Miller, R., Plessow, F., Rauh, M., Gröschl, M. and Kirschbaum, C., 2013. Comparison of salivary cortisol as measured by different immunoassays and tandem mass spectrometry. *Psychoneuroendocrinology*, 38(1), pp.50–57.

Mills, D.S., 1998. Personality and individual differences in the horse, their significance, use and measurement. *Equine Clinical Behaviour*, 30(S27), pp.10–13.

Minero, M., Zucca, D. and Canali, E., 2006. A note on reaction to novel stimulus and restraint by therapeutic riding horses. *Applied Animal Behaviour Science*, [online] 97(2–4), pp.335–342.

Momozawa, Y., Kusunose, R., Kikusui, T., Takeuchi, Y. and Mori, Y., 2005a. Assessment of equine temperament questionnaire by comparing factor structure between two separate surveys. *Applied Animal Behaviour Science*, [online] 92(1–2), pp.77–84.

Momozawa, Y., Ono, T., Sato, F., Kikusui, T., Takeuchi, Y., Mori, Y. and Kusunose, R., 2003. Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. *Applied Animal Behaviour Science*, [online] 84(84), pp.127–138.

Momozawa, Y., Takeuchi, Y., Kusunose, R., Kikusui, T. and Mori, Y., 2005b. Association between equine temperament and polymorphisms in dopamine D4 receptor gene. *Mammalian Genome*, 16(7), pp.538–544.

Momozawa, Y., Takeuchi, Y., Tozaki, T., Kikusui, T., Hasegawa, T., Raudsepp, T., Chowdhary, B.P., Kusunose, R. and Mori, Y., 2006. Polymorphism Identification, RH Mapping, and Association Analysis with the Anxiety Trait of the Equine serotonin transporter (SLC6A4) Gene. *The Journal of Veterinary Medical Science*, [online] 68(6), pp.619–621.

Monti, J.M. and Jantos, H., 2008. The roles of dopamine and serotonin, and of their receptors, in regulating sleep and waking. In: G. Di Giovanni, V. Di Matteo and E. Esposito, eds., *Progress in Brain Research*. Elsevier, pp.625–646.

Montillo, M., Comin, A., Corazzin, M., Peric, T., Faustini, M., Veronesi, M.C., Valentini, S., Bustaffa, M. and Prandi, A., 2014. The Effect of Temperature, Rainfall, and Light Conditions on Hair Cortisol Concentrations in Newborn Foals. *Journal of Equine Veterinary Science*, [online] 34(6), pp.774–778.

Montoya, E.R., Terburg, D., Bos, P.A. and van Honk, J., 2012. Testosterone, cortisol, and serotonin as key regulators of social aggression: A review and theoretical perspective. *Motivation and Emotion*, [online] 36(1), pp.65–73.

Mormède, P., Andanson, S., Aupérin, B., Beerda, B., Guémené, D., Malmkvist, J., Manteca, X., Manteuffel, G., Prunet, P., van Reenen, C.G., Richard, S. and Veissier, I., 2007. Exploration of the hypothalamic–pituitary–adrenal function as a tool to evaluate animal welfare. *Physiology & Behavior*, 92(3), pp.317–339.

Morris, P.H., Gale, A. and Duffy, K., 2002. Can judges agree on the personality of horses? *Personality and Individual Differences*, [online] 33(1), pp.67–81.

Morris, P.H., Gale, A. and Howe, S., 2002. The factor structure of horse personality. *Anthrozoos*, [online] 15(4), pp.300–322.

Mott, R., Hawthorne, S. and McBride, S.D., 2018. Spontaneous blink rate as a measure of equine stress. In: S.M. McDonnell, B. Padalino and P. Baragli, eds., *Proceedings of the 14th International Conference. Equitation Science150 years after Caprilli: theory and practice, the full circle. September 21-24, 2018*. Pisa: Pisa University Press, p.50.

Mott, R.O., Hawthorne, S.J. and McBride, S.D., 2020. Blink rate as a measure of stress and attention in the domestic horse (Equus caballus). *Scientific Reports*, [online] 10(1), pp.1–8.

Munafò, M.R., 2009. Behavioural genetics: from variance to DNA. In: P.J. Corr and G. Matthews, eds., *The Cambridge Handbook of Personality Psychology*, 1st ed. New York: Cambridge University Press, pp.287–304.

Munsters, C.C.B.M., van den Broek, J., van Weeren, R. and Sloet van Oldruitenborgh-Oosterbaan, M.M., 2013a. The effects of transport, riot control training and night patrols on the workload and stress of mounted police horses. *Applied Animal Behaviour Science*, [online] 143(1), pp.52–60.

Munsters, C.C.B.M., Visser, E.K., van den Broek, J. and Sloet van Oldruitenborgh-Oosterbaan, M.M., 2013b. Physiological and behavioral responses of horses during police training. *Animal*, [online] 7(5), pp.822–827.

Myers, J. and Myers, S., 2021. *Equiculture*. [online].

National Research Council, 2007. *Nutrient Requirements of Horses*. [online] Washington, D.C.: National Academies Press.

Neuman-Lee, L.A., Terletzky, P.A., Atwood, T.C., Gese, E.M., Smith, G.D., Greenfield, S., Pettit, J. and French, S.S., 2017. Demographic and temporal variations in immunity and condition of polar bears (Ursus maritimus) from the southern Beaufort Sea. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, [online] 327(5), pp.333–346.

Ninomiya, S., Anjiki, A., Nishide, Y., Mori, M., Deguchi, Y. and Satoh, T., 2013. Polymorphisms of the Dopamine D4 Receptor Gene in Stabled Horses are Related to Differences in Behavioral Response to Frustration. *Animals*, [online] 3(3), pp.663–669.

Noble, G.K., Blackshaw, K.L., Cowling, A., Harris, P.A. and Sillence, M.N., 2013. An objective measure of reactive behaviour in horses. *Applied Animal Behaviour Science*, [online] 144(3–4), pp.121–129.

Norton, T., Piette, D., Exadaktylos, V. and Berckmans, D., 2018. Automated real-time stress monitoring of police horses using wearable technology. *Applied Animal Behaviour Science*, 198, pp.67–74.

O'Connell, L.A. and Hofmann, H.A., 2012. Evolution of a vertebrate social decision-making network. *Science*, 336(6085), pp.1154–1157.

Odberg, F.O. and Bouissou, M.F., 1999. The development of equestrianism from the baroque period to the present day and its consequences for the welfare of horses. *Equine veterinary journal. Supplement*, (28), pp.26–30.

Oishi, Y. and Lazarus, M., 2017. The control of sleep and wakefulness by mesolimbic dopamine systems. *Neuroscience Research*, [online] 118, pp.66–73.

Olsen, H.F. and Klemetsdal, G., 2017. Temperament of the Norwegian horse breeds – a questionnaire based study. *Applied Animal Behaviour Science*, 193, pp.60–66.

Ormel, J., Bastiaansen, A., Riese, H., Bos, E.H., Servaas, M., Ellenbogen, M., Rosmalen, J.G.M. and Aleman, A., 2013. The biological and psychological basis of neuroticism: Current status and future directions. *Neuroscience & Biobehavioral Reviews*, 37(1), pp.59–72.

Oswald, L.M., Zandi, P., Nestadt, G., Potash, J.B., Kalaydjian, A.E. and Wand, G.S., 2006. Relationship between Cortisol Responses to Stress and Personality. *Neuropsychopharmacology 2006 31:7*, [online] 31(7), pp.1583–1591.

Packer, R.M.A., Davies, A.M., Volk, H.A., Puckett, H.L., Hobbs, S.L. and Fowkes, R.C., 2019. What can we learn from the hair of the dog? Complex effects of endogenous and exogenous stressors on canine hair cortisol. *PLOS ONE*, [online] 14(5), p.e0216000.

Palme, R., 2012. Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animals. *Animal Welfare*, [online] 21(3), pp.331–337.

Parker, M., Goodwin, D., Eager, R.A., Redhead, E.S. and Marlin, D.J., 2009. Comparison of Polar® heart rate interval data with simultaneously recorded ECG signals in horses. *Comparative Exercise Physiology*, [online] 6(4), pp.137–142.

Paulhus, D.L. and Vazire, S., 2007. The Self-Report Method. In: R. Robins, C. Fraley and R. Krueger, eds., *Handbook of research methods in personality psychology*. New York: The Guilford Press, pp.224–239.

Pawluski, J., Jego, P., Henry, S., Bruchet, A., Palme, R., Coste, C. and Hausberger, M., 2017. Low plasma cortisol and fecal cortisol metabolite measures as indicators of compromised welfare in domestic horses (Equus caballus). *PLOS ONE*, [online] 12(9), p.e0182257.

Pearson, G., Waran, N., Reardon, R.J.M., Keen, J. and Dwyer, C., 2021. A Delphi study to determine expert consensus on the behavioural indicators of stress in horses undergoing veterinary care. *Applied Animal Behaviour Science*, 237, p.105291.

Peeters, M., Closson, C., Beckers, J.F. and Vandenheede, M., 2013. Rider and Horse Salivary Cortisol Levels During Competition and Impact on Performance. *Journal of Equine Veterinary Science*, [online] 33(3), pp.155–160.

Peeters, M., Sulon, J., Beckers, J.-F., Ledoux, D. and Vandenheede, M., 2011a. Comparison between blood serum and salivary cortisol concentrations in horses using an adrenocorticotropic hormone challenge. *Equine Veterinary Journal*, [online] 43(4), pp.487–493.

Peeters, M., Sulon, J., Beckers, J.-F., Ledoux, D. and Vanderheede, M., 2011b. Comparison between blood serum and salivary cortisol concentrations in horses using an adrenocorticotropic hormone challenge. *Equine Veterinary Journal*, [online] 43(4), pp.487–493.

Pervin, L. and John, O.P., 1997. *Personality: Theory and research*. 7th ed. New-York: Wiley.

Phillips, A.C., Carroll, D., Burns, V.E. and Drayson, M., 2005. Neuroticism, cortisol reactivity, and antibody response to vaccination. *Psychophysiology*, [online] 42(2), pp.232–238.

Pierard, M., Hall, C., König von Borstel, U., Averis, A., Hawson, L., McLean, A., Nevison, C., Visser, K. and McGreevy, P., 2015. Evolving protocols for research in equitation science. *Journal of Veterinary Behavior: Clinical Applications and Research*, [online] 10(3), pp.255–266.

Pierard, M., McGreevy, P. and Geers, R., 2017. Developing behavioral tests to support selection of police horses. *Journal of Veterinary Behavior: Clinical Applications and Research*, 19, pp.7–13.

Pötsch, L. and Moeller, M.R., 1996. On Pathways for Small Molecules Into and Out of Human Hair Fibers. *Journal of Forensic Sciences*, 41(1), pp.121–125.

Prinsloo, M., Hynd, P., Franklin, S., Weaver, S. and van den Boom, R., 2019. Hair cortisol concentration is inversely related to the severity of equine squamous gastric disease. *Veterinary Journal*, 249, pp.58–59.

De Raad, B., 2009. Structural models of personality. In: P. Corr and G. Matthews, eds., *The Cambridge Handbook of Personality Psychology*, 1st ed. New-York: Cambridge University Press, pp.127–147.

R Core Team, 2019. *R: A Language and Environment for Statistical Computing*.

Rankins, E.M. and Wickens, C.L., 2020. A systematic review of equine personality. *Applied Animal Behaviour Science*, 231, p.105076.

Raul, J.S., Cirimele, V., Ludes, B. and Kintz, P., 2004. Detection of physiological concentrations of cortisol and cortisone in human hair. *Clinical Biochemistry*, 37(12), pp.1105–1111.

Réale, D., Reader, S.M., Sol, D., McDougall, P.T. and Dingemanse, N.J., 2007. Integrating animal temperament within ecology and evolution. *Biological Reviews*, 82(2), pp.291–318.

Ricci-Bonot, C., Romero, T., Nicol, C. and Mills, D., 2021. Social buffering in horses is influenced by context but not by the familiarity and habituation of a companion. *Scientific Reports 2021 11:1*, [online] 11(1), pp.1–10.

Rietschel, L., Streit, F., Zhu, G., McAloney, K., Frank, J., Couvy-Duchesne, B., Witt, S.H., Binz, T.M., McGrath, J., Hickie, I.B., Hansell, N.K., Wright, M.J., Gillespie, N.A., Forstner, A.J., Schulze, T.G., Wüst, S., Nöthen, M.M., Baumgartner, M.R., Walker, B.R., Crawford, A.A., Colodro-Conde, L., Medland, S.E., Martin, N.G. and Rietschel, M., 2017. Hair Cortisol in Twins: Heritability and Genetic Overlap with Psychological Variables and Stress-System Genes. *Scientific Reports*, [online] 7(1), p.15351.

Roberts, K., Hemmings, A., Moore-Colyer, M. and Hale, C., 2015. Cognitive differences in horses performing locomotor versus oral stereotypic behaviour. *Applied Animal Behaviour*

Science, [online] 168, pp.37–44.

Roberts, K., Hemmings, A.J., McBride, S.D. and Parker, M.O., 2017. Causal factors of oral versus locomotor stereotypy in the horse. *Journal of Veterinary Behavior*, [online] 20, pp.37–43.

Roberts, K., Hemmings, A.J., Moore-Colyer, M., Parker, M.O. and McBride, S.D., 2016. Neural modulators of temperament: A multivariate approach to personality trait identification in the horse ☆. *Physiology & Behavior*, [online] 167, pp.125–131.

Roebel, A.M. and MacLean, W.E., 2007. Spontaneous eye-blinking and stereotyped behavior in older persons with mental retardation. *Research in Developmental Disabilities*, [online] 28(1), pp.37–42.

Roth, L.S. V., Faresjö, Å., Theodorsson, E. and Jensen, P., 2016. Hair cortisol varies with season and lifestyle and relates to human interactions in German shepherd dogs. *Scientific Reports*, [online] 6(1), p.19631.

Rothbart, M.K., 2007. Temperament, development, and personality. *Current Directions in Psychological Science*, 16(4), pp.207–212.

Russell, E., Koren, G., Rieder, M. and Van Uum, S., 2012. Hair cortisol as a biological marker of chronic stress: Current status, future directions and unanswered questions. *Psychoneuroendocrinology*, [online] 37(5), pp.589–601.

Safryghin, A., Hebesberger, D. V. and Wascher, C.A.F., 2019. Testing for Behavioral and Physiological Responses of Domestic Horses (Equus caballus) Across Different Contexts – Consistency Over Time and Effects of Context. *Frontiers in Psychology*, [online] 10(APR), p.849.

Salaberger, T., Millard, M., Makarem, S.E.S. El, Möstl, E., Grünberger, V., Krametter-Frötscher, R., Wittek, T. and Palme, R., 2016. Influence of external factors on hair cortisol concentrations. *General and Comparative Endocrinology*, [online] 233, pp.73–78.

Salimetrics, 2019. *High Sensitivity SALIVARY CORTISOL enzyme immunoassay kit*. [online] salimetrics.com.

Salimetrics, 2020. *Calculating Inter- and Intra-Assay Coefficients of Variability*. [online] salimetrics.com.

Sapolsky, R.M., Romero, L.M. and Munck, A.U., 2000. How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocrine Reviews*, [online] 21(1), pp.55–89.

Sauer, F.J., Gerber, V., Frei, S., Bruckmaier, R.M. and Groessl, M., 2020. Salivary cortisol measurement in horses: immunoassay or LC-MS/MS? *Domestic Animal Endocrinology*, 72, p.106445.

Sauer, F.J., Hermann, M., Ramseyer, A., Burger, D., Riemer, S. and Gerber, V., 2019. Effects of breed, management and personality on cortisol reactivity in sport horses. *PLoS ONE*, [online] 14(12), p.e0221794.

Sauveroche, M., Henriksson, J., Theodorsson, E., Svensson Holm, A.C. and Roth, L.S.V., 2020. Hair cortisol in horses (Equus caballus) in relation to management regimes, personality, and breed. *Journal of Veterinary Behavior*, 37, pp.1–7.

Scandell, D.J., 2000. Development and initial validation of validity scales for the NEO-Five factor inventory. *Personality and Individual Differences*, 29(6), pp.1153–1162.

Scheidegger, M.D., Gerber, V., Ramseyer, A., Schüpbach-Regula, G., Bruckmaier, R.M. and van der Kolk, J.H., 2016. Repeatability of the ACTH stimulation test as reflected by salivary cortisol response in healthy horses. *Domestic Animal Endocrinology*, 57, pp.43–47.

Schell, C.J., Young, J.K., Lonsdorf, E.V., Mateo, J.M. and Santymire, R.M., 2017. Investigation of techniques to measure cortisol and testosterone concentrations in coyote hair. *Zoo Biology*, 36(3), pp.220–225.

Schmidt, A., Aurich, J., Möstl, E., Müller, J. and Aurich, C., 2010a. Changes in cortisol release and heart rate and heart rate variability during the initial training of 3-year-old sport horses. *Hormones and Behavior*, [online] 58(4), pp.628–636.

Schmidt, A., Biau, S., Möstl, E., Becker-Birck, M., Morillon, B., Aurich, J., Faure, J.-M. and Aurich, C., 2010b. Changes in cortisol release and heart rate variability in sport horses during long-distance road transport. *Domestic Animal Endocrinology*, [online] 38(3), pp.179–189.

Schmidt, A., Hödl, S., Möstl, E., Aurich, J., Müller, J. and Aurich, C., 2010c. Cortisol release, heart rate, and heart rate variability in transport-naive horses during repeated road transport. *Domestic Animal Endocrinology*, [online] 39(3), pp.205–213.

Schmidt, A., Möstl, E., Wehnert, C., Aurich, J., Müller, J. and Aurich, C., 2010d. Cortisol release and heart rate variability in horses during road transport. *Hormones and Behavior*, [online] 57(2), pp.209–215.

Schubach, K.M., Cooke, R.F., Brandão, A.P., Lippolis, K.D., Silva, L.G.T., Marques, R.S. and Bohnert, D.W., 2017. Impacts of stocking density on development and puberty attainment of replacement beef heifers. *animal*, [online] 11(12), pp.2260–2267.

Le Scolan, N., Hausberger, M. and Wolff, A., 1997. Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. *Behavioural Processes*, [online] 41(3), pp.257–266.

Seaman, S.C., Davidson, H.P.B. and Waran, N.K., 2002. How reliable is temperament assessment in the domestic horse (Equus caballus)? *Applied Animal Behaviour Science*, [online] 78(2–4), pp.175–191.

Sgoifo, A., Carnevali, L. and Grippo, A.J., 2014. The socially stressed heart. Insights from studies in rodents. *Neuroscience and Biobehavioral Reviews*, [online] 39, pp.51–60.

Shanahan, S., 2003. Trailer Loading Stress in Horses: Behavioral and Physiological Effects of Nonaversive Training (TTEAM). *Journal of Applied Animal Welfare Science*, [online] 6(4), pp.263–274.

Shimohata, A., Ishihara, K., Hattori, S., Miyamoto, H., Morishita, H., Ornthanalai, G., Raveau, M., Ebrahim, A.S., Amano, K., Yamada, K., Sago, H., Akiba, S., Mataga, N., Murphy, N.P., Miyakawa, T. and Yamakawa, K., 2017. Ts1Cje Down syndrome model mice exhibit environmental stimuli-triggered locomotor hyperactivity and sociability concurrent with increased flux through central dopamine and serotonin metabolism. *Experimental Neurology*, 293, pp.1–12.

Shrout, P.E. and Fleiss, J.L., 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, [online] 86(2), pp.420–428.

Sigurjónsdóttir, H., Van Dierendonck, M.C., Snorrason, S. and Thórhallsdóttir, A.G., 2003. Social Relationships in a Group of Horses without a Mature Stallion. *Behaviour*, [online] 140(6), pp.783–804.

Sih, A. and Del Giudice, M., 2012. Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, [online] 367(1603), pp.2762–72.

Simms, L.J. and Watson, D., 2007. The Construct Validation Approach to Personality Scale Construction. In: R.W. Robins, R.C. Fraley and R.F. Krueger, eds., *Handbook of research methods in personality psychology*. New York: The Guilford Press, pp.240–258.

Siniscalchi, M., McFarlane, J.R., Kauter, K.G., Quaranta, A. and Rogers, L.J., 2013. Cortisol levels in hair reflect behavioural reactivity of dogs to acoustic stimuli. *Research in Veterinary Science*, 94(1), pp.49–54.

Slagter, H.A., Georgopoulou, K. and Frank, M.J., 2015. Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia*, 71, pp.126–132.

Sneddon, L.U., 2003. The bold and the shy: Individual differences in rainbow trout. *Journal of Fish Biology*, 62(4), pp.971–975.

Sotohira, Y., Suzuki, K., Sano, T., Arai, C., Asakawa, M., Hayashi, H., Siziku, K., Sano, T., Arai, C., Asakawa, M. and Hayashi, H., 2017. Stress assessment using hair cortisol of kangaroos affected by the lumpy jaw disease. *Journal of Veterinary Medical Science*, [online] 79(5), pp.852–854.

Squibb, K., Griffin, K., Favier, R. and Ijichi, C., 2018. Poker Face: Discrepancies in behaviour and affective states in horses during stressful handling procedures. *Applied Animal Behaviour Science*, [online] 202, pp.34–38.

Starling, M., McLean, A. and McGreevy, P., 2016. The Contribution of Equitation Science to Minimising Horse-Related Risks to Humans. *Animals*, [online] 6(3), p.15.

Steptoe, A., Easterlin, E. and Kirschbaum, C., 2017. Conscientiousness, hair cortisol concentration, and health behaviour in older men and women. *Psychoneuroendocrinology*, [online] 86, pp.122–127.

Stevenson-Hinde, J., Stillwell-Barnes, R. and Zunz, M., 1980. Subjective assessment of rhesus monkeys over four successive years. *Primates*, [online] 21(1), pp.66–82.

Stevenson-Hinde, J. and Zunz, M., 1978. Subjective Assessment of Individual Rhesus Monkeys. *Primates*, 19(3), p.473.

Stradaioli, G., Peric, T., Montillo, M., Comin, A., Corazzin, M., Veronesi, M.C.M.C. and Prandi, A., 2017. Hair cortisol and testosterone concentrations and semen production of Bos taurus bulls. *Italian Journal of Animal Science*, [online] 16(4), pp.631–639.

Stucke, D., Ruse, M.G., Lebelt, D., Große Ruse, M. and Lebelt, D., 2015. Measuring heart rate variability in horses to investigate the autonomic nervous system activity – Pros and cons of different methods. *Applied Animal Behaviour Science*, [online] 166(1), pp.1–10.

Suwała, M., Górecka-Bruzda, A., Walczak, M., Ensminger, J. and Jezierski, T., 2016. A desired profile of horse personality – A survey study of Polish equestrians based on a new approach to equine temperament and character. *Applied Animal Behaviour Science*, [online] 180, pp.65–77.

Task Forceof the European Society of Cardiology and the North American Society of Pacing Electrophysiology, 1996. Heart Rate Variability. *Circulation*, [online] 93(5), pp.1043–1065.

Taylor, K.D. and Mills, D.S., 2006. The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior*, 1(3), pp.94–108.

Tharp, I.J. and Pickering, A.D., 2011. Individual differences in cognitive-flexibility: The influence of spontaneous eyeblink rate, trait psychoticism and working memory on attentional set-shifting. *Brain and Cognition*, 75(2), pp.119–125.

Unsworth, N., Robison, M.K. and Miller, A.L., 2019. Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective and Behavioral Neuroscience*, [online] 19(4), pp.1074–1093.

Valenchon, M., Levy, F., Gorecka-Bruzda, A., Calandreau, L., Lansade, L.L., Lévy, F., Górecka-Bruzda, A., Calandreau, L. and Lansade, L.L., 2013. Characterization of long-term memory, resistance to extinction, and influence of temperament during two instrumental tasks in horses. *Animal Cognition*, [online] 16(6), pp.1001–1006.

Vazire, S., Gosling, S.D., Dickey, A.S. and Schapiro, S.J., 2007. Measuring Personality in Nonhuman Animals. In: R.W. Robins, R.C. Fraley and R.F. Krueger, eds., *Handbook of research* *methods in personality psychology*. New York: The Guilford Press, pp.190–206.

Veenema, A.H., Meijer, O.C., de Kloet, E.R., Koolhaas, J.M. and Bohus, B.G., 2003. Differences in basal and stress-induced HPA regulation of wild house mice selected for high and low aggression. *Hormones and Behavior*, [online] 43(1), pp.197–204.

Villas-Boas, J.D., Dias, D.P.M., Trigo, P.I., Almeida, N.A. dos S., Almeida, F.Q. de and Medeiros, M.A. de, 2016. Behavioural, endocrine and cardiac autonomic responses to a model of startle in horses. *Applied Animal Behaviour Science*, [online] 174, pp.76–82.

Visser, E., Van Reenen, C., Engel, B., Schilder, M., Barneveld, A. and Blokhuis, H., 2003a. The association between performance in show-jumping and personality traits earlier in life. *Applied Animal Behaviour Science*, [online] 82(4), pp.279–295.

Visser, E., van Reenen, C., van der Werf, J., Schilder, M., Knaap, J., Barneveld, A. and Blokhuis, H., 2002. Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiology & Behavior*, [online] 76(2), pp.289–296.

Visser, E., Van Reenen, C., Zetterqvist Blokhuis, M., Morgan, E., Hassmén, P., Rundgren, T. and Blokhuis, H.J., 2008. Does Horse Temperament Influence Horse–Rider Cooperation? *Journal of Applied Animal Welfare Science*, [online] 11, pp.267–284.

Visser, E., Van Reenen, C.G., Rundgren, M., Zetterqvist, M., Morgan, K. and Blokhuis, H.J., 2003b. Responses of horses in behavioural tests correlate with temperament assessed by riders. *Equine veterinary journal*, 35(2), pp.176–183.

Volpe, G., Draisci, R., Palleschi, G. and Compagnone, D., 1998. 3,3′,5,5′-Tetramethylbenzidine as electrochemical substrate for horseradish peroxidase based enzyme immunoassays. A comparative study. *The Analyst*, [online] 123(6), pp.1303–1307.

Walton, A. and Toth, A.L., 2016. Variation in individual worker honey bee behavior shows hallmarks of personality. *Behavioral Ecology and Sociobiology 2016 70:7*, [online] 70(7), pp.999–1010.

Wathan, J., Burrows, A.M., Waller, B.M. and McComb, K., 2015. EquiFACS: The Equine Facial Action Coding System. *PLOS ONE*, [online] 10(9), p.e0131738.

Wennig, R., 2000. Potential problems with the interpretation of hair analysis results. *Forensic Science International*, 107(1–3), pp.5–12.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H. and RStudio, 2019. *Create Elegant Data Visualisations Using the Grammar of Graphics*.

Wipper, A., 2000. The Partnership: The Horse-Rider Relationship in Eventing. *Symbolic Interaction*, 23(1), pp.47–70.

Wolff, A., Hausberger, M. and Le Scolan, N., 1997. Experimental tests to assess emotionality in horses. *Behavioural Processes*, [online] 40(3), pp.209–221.

Wolframm, I.A., Gerardus, R. and Meulenbroek, J., 2012. Co-variations between perceived personality traits and quality of the interaction between female riders and horses. *Applied Animal Behaviour Science*, [online] 139, pp.96–104.

Wood, S. and Scheipl, F., 2017. *Generalized Additive Mixed Models using 'mgcv' and 'lme4'*.

Wu, Z., Begley, C.G., Situ, P., Simpson, T. and Liu, H., 2014. The Effects of Mild Ocular Surface Stimulation and Concentration on Spontaneous Blink Parameters. *Current Eye Research*, [online] 39(1), pp.9–20.

XiuJuan, R., Hong, Y., YiPing, Z., Sarula, ShaoFeng, S., XiSheng, W., HongMei, B., DongYi, B., Bei, L., Wunierfu, S. and Manglai, D., 2015. Association analysis between major temperament traits and diversification of the candidate gene in Mongolian horse (Equus caballus). *Journal of Agricultural Biotechnology*, [online] 25(3), pp.405–414.

Yamanashi, Y., Morimura, N., Mori, Y., Hayashi, M. and Suzuki, J., 2013. Cortisol analysis of hair of captive chimpanzees (Pan troglodytes). *General and Comparative Endocrinology*, [online] 194, pp.55–63.

Yamanashi, Y., Teramoto, M., Morimura, N., Hirata, S., Inoue-Murayama, M. and Idani, G., 2016a. Effects of Relocation and Individual and Environmental Factors on the Long-Term Stress Levels in Captive Chimpanzees (Pan troglodytes): Monitoring Hair Cortisol and Behaviors. *PLOS ONE*, [online] 11(7), p.e0160029.

Yamanashi, Y., Teramoto, M., Morimura, N., Hirata, S., Suzuki, J., Hayashi, M., Kinoshita, K., Murayama, M. and Idani, G., 2016b. Analysis of hair cortisol levels in captive chimpanzees: Effect of various methods on cortisol stability and variability. *MethodsX*, [online] 3, pp.110–117.

Yamanashi, Y., Teramoto, M., Morimura, N., Nogami, E. and Hirata, S., 2018. Social relationship and hair cortisol level in captive male chimpanzees (Pan troglodytes). *Primates*, [online] 59(2), pp.145–152.

Yarnell, K., Hall, C. and Billett, E., 2013. An assessment of the aversive nature of an animal management procedure (clipping) using behavioral and physiological measures. *Physiology & Behavior*, [online] 118, pp.32–39.

Yarnell, K., Hall, C., Royle, C. and Walker, S.L., 2015. Domesticated horses differ in their behavioural and physiological responses to isolated and group housing. *Physiology & Behavior*, [online] 143, pp.51–57.

Zaman, M.L. and Doughty, M.J., 1997. Some methodological issues in the assessment of the spontaneous eyeblink frequency in man. *Ophthalmic and Physiological Optics*, [online] 17(5), pp.421–432.

Zilioli, S., Ponzi, D., Henry, A. and Maestripieri, D., 2015. Testosterone, Cortisol and Empathy: Evidence for the Dual-Hormone Hypothesis. *Adaptive Human Behavior and Physiology*, [online] 1(4), pp.421–433.

Zillig, L.M.P., Hemenover, S.H. and Dienstbier, R.A., 2002. What Do We Assess when We Assess a Big 5 Trait? A Content Analysis of the Affective, Behavioral, and Cognitive Processes Represented in Big 5 Personality Inventories. *Personality and Social Psychology Bulletin*, [online] 28(6), pp.847–858.

Zuur, A.F., Ieno, E.N. and Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, [online] 1(1), pp.3–14.

Appendix A. Consistency of autonomic and HPA axis responses across test situations

A. METHODS

i. Summary of physiological variables and test procedures

Horses (n = 23; 15 geldings, 8 mares; mean age 11.7±4.0 years) were exposed to three test situations expected to induce short-term mild stress responses and commonly used in behaviour-based personality assessment: a Novel Object test, a Startle test and a handling test (Sham Clipping). The test procedures followed a 5 minute period of habituation to the test arena (Novel Object and Startle tests) or baseline physiological monitoring in the home stable (Sham Clipping). More details of the test procedures are available in Section 4.2.2.2 - Test procedures.

For each test situation, mean heart rate (HR), RMSSD and salivary cortisol responses to the tests were recorded. More details and justifications of the methods used to collect and process physiological data are available in Section 4.2.2.3 for mean heart rate and RMSSD, and in Section 4.2.2.4 for salivary cortisol responses.

For all three tests, mean HR and RMSSD were recorded during the 5 minute habituation or baseline phase and during the 5 minute test phase. The mean HR response to the test situation (ΔHR) was then defined as the difference between mean HR during the test and mean HR during habituation/baseline. Similarly, the RMSSD response to the test situation (ΔRMSSD) was defined as the difference between RMSSD during the test and RMSSD during habituation/baseline. For all three tests, baseline and peak salivary cortisol concentrations were also determined, defined respectively as the concentration immediately before the test and the maximum concentration post-test. The salivary cortisol response to the test (ΔCORT) was then defined as the difference between peak salivary cortisol concentration post-test and salivary cortisol concentration at baseline.

ii. Statistical analysis

All statistical analysis was carried out using R (version 3.6.1; R Core Team, 2019). The consistency of physiological responses (ΔHR, ΔRMSSD and ΔCORT) across test situations was tested using Spearman rank correlations. The Benjamini-Hochberg procedure was used to minimize the number of Type I errors associated with multiple testing, with the false discovery rate fixed at 10% (Benjamini and Hochberg, 1995). For each correlation, both uncorrected (p) and adjusted (adj. p) p values are reported.

B. RESULTS

There were no significant correlations between any of the HR responses to the different test situations, either before or after a Benjamini-Hochberg correction for multiple testing (false discovery rate: 10%) was applied (Table A).

There were no significant correlations between any of the RMSSD responses to the different test situations, either before or after a Benjamini-Hochberg correction for multiple testing (false discovery rate: 10%) was applied (Table A).

There were no significant correlations between any of the HR responses to the different test situations, either before or after a Benjamini-Hochberg correction for multiple testing (false discovery rate: 10%) was applied (Table A).

Table A – Summary of Spearman rank correlations between physiological changes in response to each of the three test situations. Both original and Benjamini-Hochberg adjusted (false discovery rate = 10%) p values are presented.

	Test 1	Test 2	N	r	p value	Adjusted p value
ΔHR	Novel Object	Startle	23	-0.005	0.982	0.982
	Novel Object	Sham Clipping	20	-0.128	0.601	0.902
	Startle	Sham Clipping	20	0.237	0.327	0.902
ΔRMSSD	Novel Object	Startle	23	0.314	0.177	0.266
	Novel Object	Sham Clipping	20	0.184	0.422	0.422
	Startle	Sham Clipping	20	0.344	0.163	0.266
ΔCORT	Novel Object	Startle	23	-0.274	0.229	0.292
	Novel Object	Sham Clipping	20	-0.248	0.292	0.292
	Startle	Sham Clipping	20	0.388	0.112	0.292

C. DISCUSSION

Spearman correlation analyses suggest that the three physiological responses considered here were not consistent across test situations. Therefore, HPA axis and autonomic reactivity differed as a function of the test situation horses were exposed to, and rank order within the cohort was not conserved across situations. This is unexpected, as both HPA axis and autonomic reactivity are generally considered a trait (Koolhaas et al., 2010) linked with individual differences in underlying neuroendocrine systems (de Boer, Buwalda and Koolhaas, 2017). In horses, mean HR during behavioural tests has previously been found to be consistent across time (Visser et al., 2002; Safryghin, Hebesberger and Wascher, 2019) and across situations (König von Borstel et al., 2011; Safryghin, Hebesberger and Wascher, 2019; Mccall et al., 2006; Christensen, Keeling and Nielsen, 2005), although some contradictory evidence is available (Lansade, Bouissou and Erhard, 2008). In addition, RMSSD during behavioural tests has also been found to be consistent across time (Visser et al., 2002). To the best of the author’s knowledge, the consistency of cortisol responses across time and situations has not been studied in the horse. However, cortisol responses to an ACTH challenge are consistent across time in the horse (Scheidegger et al., 2016). The results of this work therefore contradict existing literature and do not align with the hypothesis that physiological responses to stressors constitute a trait in domestic horses.

It is possible that HPA axis and autonomic responses of the horses in the sample differed across situation due to the impact of previous experiences. Indeed, habituation has been shown to dampen heart rate responses towards the target object or procedure, but does not affect responses towards sufficiently different stimuli (Leiner and Fendt, 2011; Christensen, Zharkikh and Ladewig, 2008). Because all three test situations differed significantly, it is likely that horses in the sample may have had differing levels of previous experience with each situation, and that they would be unlikely to generalise any habituation to a given situation to the others. However, it should also be noted that all studies cited above compared the values of the physiological variable itself in the test, rather than the change observed. This may also contribute to explain the contradictory findings reported here. Indeed, in humans, evidence suggests that peak salivary cortisol and peak salivary cortisol reactivity are two independent measures, with the first reflecting the total unbound cortisol production of an individual, while the second describes the change in cortisol levels in response to the experimental situation (Khoury et al., 2015). Therefore, while the literature suggests that the absolute values of physiological variables may be consistent across tests, more work might be needed to establish whether those measures truly represent reactivity.

Appendix B. Preliminary studies for blink rate analysis

The aim of this pilot study was to inform the methods used for analysis in Chapter 6: Striatal dopamine as a potential correlate of equine personality. All pilot analyses were carried out on sections of footage collected for Chapter 6. Full details of data collection and justification of methodological choices are available in Section 6.2.2.1. Briefly, thirty minutes long video recordings of undisturbed, loosely tethered horses' face and eyes were obtained on three consecutive days, using tripod-mounted GoPro Hero 7s set to record continuously at 1440 pixels/60 frames per second in the Wide FOV mode. Ahead of blink counts, 5-minute sections of footage suitable for analysis (see Section 6.2.2.2) were selected. All sections of footage selected for blink count were zoomed in and cropped to centre on the eye.

For the purpose of both this pilot study and the final analysis carried out in 144Chapter 6, blinks were defined as "any appreciable downward movement of the upper eyelid to cover some or all of the corneal surface, immediately followed by eye reopening". A justification of this definition is available in 6.2.2.3: Blink count. The total number of blinks recorded therefore included full blinks (Roberts et al., 2016; Best et al., 2018; Merkies et al., 2019), but also partial blinks (Wathan et al., 2015; Best et al., 2018) and eyelid twitches (Merkies et al., 2019).

A. INTRA-RATER RELIABILITY OF BLINK COUNTS

Equine studies assessing SBR retrospectively from video footage typically use a single blink count (Best et al., 2018; Cherry et al., 2020; Merkies et al., 2019; Mott, Hawthorne and McBride, 2020). However, counting errors on the experimenter's part have been identified as a potential source of unreliability in blink counts (Zaman and Doughty, 1997). Therefore, to ensure reliable blink rate data was produced, the consistency of repeated counts by the same experimenter was evaluated.

A. METHODS

i. Footage selection

5-minutes segments of footage suitable for analysis (n=11) taken from 2 horses were selected at random for inclusion in the intra-rater reliability analysis of blink counts.

ii. Blink count

Videos were played at speed 0.50x on VLC Media Player and blinks were counted as they occurred using an online tally counter. The total number of blinks in each video was recounted three times by the same experimenters, in three separate viewings. In order to limit any bias for the intra-rater analysis, the experimenter was kept blind to the count being accrued by masking the display window of the online tally counter until the end of the video had been reached.

iii. Statistical analysis

Statistical analyses for this pilot study were performed using R version 3.6.1 (R Core Team, 2019). The consistency with which the experimenter counted blinks in videos was evaluated using intra-rater reliability analysis. This was done in order to determine whether an accurate assessment of the number of blinks in a video could be obtained in a single count, or whether repeated counts would be necessary to account for discrepancies in blink identification. The total number of blinks obtained on each of three recounts of 5-minute videos (n=11) were compared using intra-class correlation (ICC) analysis. Model selection for this intra-rater reliability analysis was based on Koo and Li (2016)'s decision tree and a single-ratings (k=3), absolute agreement, two-way mixed effects model (McGraw and Wong, 1996) was used. Analysis was carried out in R using the ***icc*** function from the ***irr*** package (Gamer, Lemon and Fellows Puspendra Singh, 2019), generating an ICC coefficient and 95% confidence interval. Interpretation in terms of reliability of the blink counts was carried out using the thresholds for poor (ICC<0.5), moderate (0.5<ICC<0.75), good (0.75<ICC<0.9) and excellent (ICC>0.9) agreement proposed by Koo and Li (2016).

B. RESULTS

ICC analysis of the three repeated blink counts resulted in an intra-class correlation coefficient of ICC=0.99. The 95% confidence interval was 0.974 < ICC < 0.997.

C. CONCLUSIONS

According to the threshold proposed by Koo and Li (2016), the ICC coefficient and confidence interval indicated excellent consistency between the three counts (ICC>0.9). Therefore, although counting errors on the experimenter's part have been identified as a potential source of unreliability in blink counts (Zaman and Doughty, 1997), here intra-rater reliability was high. For this reason, a single count rather than 3 recounts may be used for data processing in Chapter 6: Striatal dopamine as a potential correlate of equine personality, with minimal impact of the reliability of blink counts.

B. IMPACT OF WAKEFULNESS STATE ON SBR

Informal observations during footage selection suggested that wakefulness state may have an important impact on SBR. To the author's knowledge, this effect has not yet been described in the equine literature; however, fatigue and drowsiness are known to impact blink characteristics in humans (Cori et al., 2019; Jongkees and Colzato, 2016). Thus, wakefulness state might act as a confounding factor of SBR in the horse and may need to be taken into consideration on analyses of SBR. A pilot study was therefore carried out to formally investigate whether wakefulness state impacted blink rate in the sample of horses used in this project.

A. METHODS

i. Footage selection

To investigate the impact of wakefulness state on SBR, 30-minute videos containing footage of the horse in both an active and a dozing state were used (n=11). "Dozing" was identified based on previously published ethograms as a behavioural state in which the horse stands inactive with the head lowered, the lower lip relaxed and eyes partially closed; muscle twitches and leaning behaviour may be observed or weight may be borne on three legs (McDonnell, 2003). "Asleep with eyes closed" was also identified, as a state sharing the characteristics of Dozing but in which the corneal surface was fully covered by the eyelids for more than a second. Any state in which the horse was not “Dozing” and “Asleep with eyes closed” was recorded as “Active”. SBR in an active state was evaluated by counting blinks in a randomly selected 5-minutes segment of the video in which the horse was active, while SBR while dozing was evaluated by counting blinks in a randomly selected 5-minutes segment of the same video in which the horse was dozing. This resulted in a dataset of 11 paired observations.

ii. Blink count

Videos were played at speed 0.50x on VLC Media Player and blinks were counted as they occurred using an online tally counter. Given the high intra-rater reliability of blink counts evidenced in Appendix B, section A, the total number of blinks was only counted once for each paired video.

iii. Statistical analysis

Statistical analyses for this pilot study were performed using R version 3.6.1 (R Core Team, 2019). The impact of dozing behaviour on SBR was evaluated by comparing mean SBR in paired videos from horses (n=7) while active and dozing. Residuals from the paired observations were tested for normality using a Shapiro-Wilk test, appropriate for small sample sizes (n<50). Following confirmation of normality, observations from the two groups were tested for homogeneity of variance using Levene’s Test. A paired T-test was then used to compare mean Spontaneous Eye Blink Rate in the two states.

B. RESULTS

Wakefulness state had a highly significant impact on SBR (paired T-test: n=11; t_{10}=5.94; p=0.00014; Cohen's d=1.88). SEBR was significantly higher (mean ± SD: 17.59 ± 4.07 blinks/min) when horses were active than when they were drowsing (mean ± SD: 9.29 ± 3.36 blinks/min) (Figure A).

C. CONCLUSIONS

SBR was significantly lower in videos showing horses in a dozing state compared to an active state. This demonstrates a clear impact of wakefulness state on SBR, in accordance with informal observations and findings in humans (Cori et al., 2019; Jongkees and Colzato, 2016). Wakefulness state therefore acts as a confounding factor of SBR and should be controlled for in studies utilising SBR as an outcome variable.

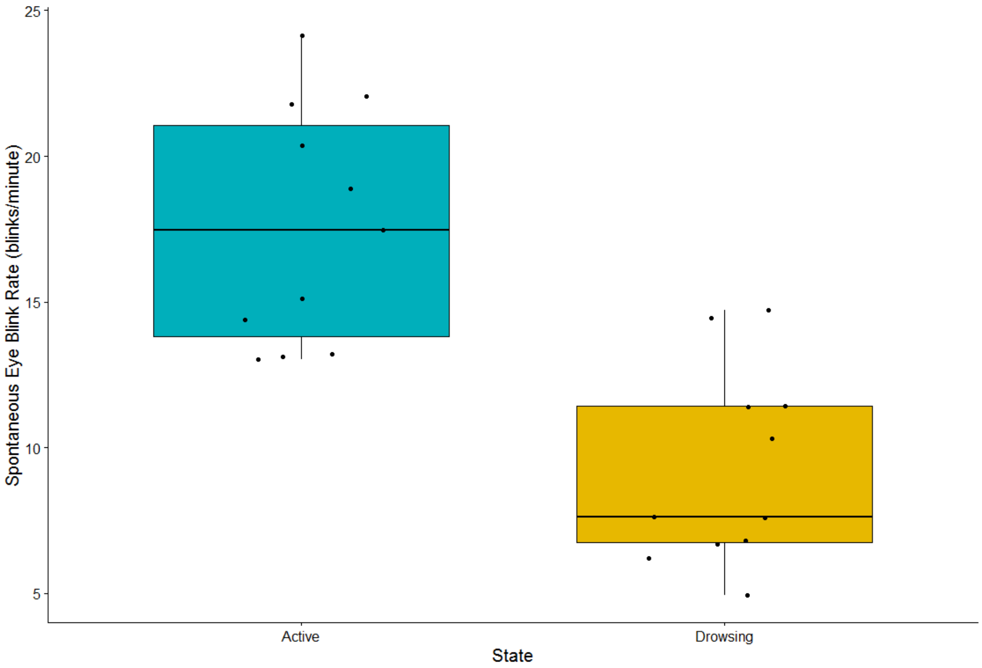

Figure A – Boxplots comparing the spontaneous eye blink rate (blinks/minute) of horses during periods of activity compared to periods of drowsing.

C. CONSISTENCY OF SBR OVER THREE DAYS

This pilot analysis was carried out on the final blink counts obtained from the footage selected for blink counts in Chapter 6 (see Sections 6.2.2.2 and 6.2.2.3). Briefly, SBR was determined from 5-minute sections of footage collected on 3 consecutive days in horses (n=20) loosely tied in their home stables. This pilot analysis aimed to evaluate whether SBR was consistent over the three days, in order to determine whether repeated measures could meaningfully be averaged ahead of the final analysis

A. METHODS

The repeatability of SBR over the three days was evaluated using ICC. A single-ratings (k=3), absolute agreement, two-way mixed effects model was used (Koo and Li, 2016) and analysis was carried out in R using the ***icc*** function from the ***irr*** package (Gamer, Lemon and Fellows Puspendra Singh, 2019), generating an ICC coefficient and 95% confidence interval.

B. RESULTS

ICC analysis of SBR on the 3 consecutive days of data collection resulted in an intra-class correlation coefficient of ICC=0.157. The 95% confidence interval was -0.101 < ICC < 0.472. According to the threshold proposed by Koo and Li (2016), this indicated very poor repeatability of SBR over the three days (ICC < 0.5).

C. DISCUSSION AND CONCLUSION

This pilot analysis revealed that the repeated measures of SBR had very low repeatability over the three days of data collection. One potential source of unreliability in blink counts is counting errors on the experimenter's part (Zaman and Doughty, 1997). However, this is unlikely to have driven unreliability here: footage where the eye was difficult to observe was excluded from analysis, and intra-rater reliability analysis of the blink counts showed that counts were highly consistent (see Appendix B, Section A). However, in humans SBR has also been shown to be very susceptible to phasic changes in response to distractions (Březinová and Kendell, 1977; Bacher and Allen, 2009; Doughty, 2001) or changes in experimental conditions (Doughty, 2016). Here, it is likely that a number of confounding factors of SBR may have affected consistency. In particular, SBR has been shown to change rapidly in response to attentional processes in the horse (Mott, Hawthorne and McBride, 2020; Merkies et al., 2019). In addition, the present work demonstrated a strong phasic influence of wakefulness state on SBR (see Appendix B, Section B). Therefore, it is possible that the low repeatability observed was driven in part by attentional processes as the horses responded to environmental distractions. To the author's knowledge this is the first study investigating the repeatability of SBR measured at rest in the horse. This result suggests that more research is needed to help inform protocol designs used to measure tonic SBR.

In the context of the analysis carried out in Chapter 6, the very low consistency of SBR over the three days of data collection meants that the three repeated measures could not meaningfully be averaged.