scientific reports

Check for updates

OPEN

Exploring the induction and measurement of positive affective state in equines through a personality-centred lens

Loni Loftus[1,2,3✉], Amy Newman[3], Matthew Leach[4] & Lucy Asher[1]

There is increasing focus on how to induce and measure positive affective states in animals and the development of social license to operate has brought this to the forefront within equestrianism. This study aimed to utilise a range of methods to induce and measure positive affect in horses in real-world settings. Twenty healthy horses were scored for personality, exposed to four induction methods (wither scratching, high value food provision, positive reinforcement training and the addition of an affiliative conspecific), and data collected on their behaviour (QBA and ethograms) and physiology (heart and respiratory rate, heart rate variability, eye and ear thermography and salivary cortisol). Analyses identified potentially sensitive and specific behavioural (ear and eye position, QBA items, frustration items) and physiological (RR mean, HF power, LF power, LF/HF ratio, mean HR, RMSSD and pNN50) measures of affective state across the four quadrants of core affect. Individual difference effects were found, and personality traits such as unfriendly, nervous and unresponsive were associated with differing responses to induction stimuli indicating that all four induction stimuli are potentially useful for inducing positive affect depending on their salience to the individual. Research measuring and inducing positive affect in animals rarely considers personality, but this study underscores its importance. The dimensional approach taken allowed for assessment of the broad arousal and valence components of affect without ascribing measures to discrete emotions. Accurate, real-world measures of affect could benefit 116 million equines globally, and exploring ways to promote positive affect in horses can significantly enhance their welfare.

Emotions (short-term valenced states)[1] are notoriously difficult to study in animals since they are not able to self-report. However, over the last twenty years research has progressed our understanding of how to assess animal emotion[2,3]. 'Affect' is an overarching term for states (affective states) that have valence (they are positive or negative). Affect includes short-term emotions and longer-term moods but also the valenced components of sensations and non-conscious components such as behavioural indicators and neural changes[1]. Core affect refers to the change in the internal environment in two key dimensions; emotional valence (positive and negative) and arousal (a state of physiological alertness and readiness for action)[1,4]. A rise in welfare standards, coupled with the requirement to consider both physiological and psychological welfare within welfare assessment, has led to a renewed focus on emotions in animals and positive welfare states including positive affective states[5]. The development of Social Licence to Operate (SLO) (societal approval or acceptance of the activity) in the equestrian industry further motivates research focussed on promoting positive affect in domesticated equines to enhance welfare[6,7].

Behavioural and physiological measures are used as proxies for identification of affective state in animals. These outputs can be considered in terms of opportunities for pleasure (positive affect) or avoidance of threats (negative affect)[8]. From a behavioural perspective, affiliative behaviour, play and information gathering activities have been proposed as potential positive indicators[9]. These are behaviours said to be mediated by the seeking,

[1]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. [2]Royal (Dick) School of Veterinary Studies and the Roslin Institute, Easter Bush Campus, University of Edinburgh, Midlothian EH25 9RG, UK. [3]University Centre Askham Bryan, Askham Bryan, York YO23 3FR, UK. [4]Comparative Biology Centre, Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. ✉email: i.l.loftus@newcastle.ac.uk; loni.loftus@ed.ac.uk

care, and play systems[10] and therefore unlikely to be performed by animals in negative affect[1,2,11,12]. Some of these are included in behavioural measures of the AWIN project[13] which developed a welfare assessment protocol for horses which included: social interaction; stereotypy performance (or absence of); fear (or absence of); human-animal relationships assessed via avoidance distance, voluntary animal approach and forced human approach tests; and qualitative behaviour assessment (QBA). QBA was included in the 2004–2009 European Commission's Welfare Quality Audit and identified as the only measure which captures positive welfare[14] however, some authors have reported lower levels of reliability when using QBA as a standalone welfare assessment tool[15]. Hall and Heleski[16] suggest that ethograms of behaviour are important means of evaluating both the affective and physical welfare of horses alongside physiological measures. Whilst overt body language is important in understanding horse welfare, more subtle signs may be displayed by equines either prior to, or in place of, more overt body language[17]; each can be evaluated via ethograms. Facial actions can be reliably recorded using the Equine Facial Action Coding System, EquiFACS[18], which has previously been utilised in attempts to study positive affective state[19–21]. Subtle indicators might also include lateralisation of behaviour, which has been more widely studied in other species (reviewed by Rogers et al.[22]) and recently in horses[23,24]. One further aspect of behaviour that is relevant to the current study is frustration, defined as the thwarting of impulses or actions that prevents individuals from obtaining something they expect based on past experience[25,26]. In studies presenting affective stimuli there is potential for transient alteration in affective state, for example where initial positive affect may switch to negative affective states, such as frustration, if the goal is not achieved[27]. Frustration may be evaluated using previously identified behavioural signs, including repetitive behaviour[28].

Physiological measures have also been used as indicators of affective state in horses. Heart rate and heart rate variability (HRV, using time-, frequency- and non-linear-domains) have been widely used in the assessment of stress in horses[29–32]. Whilst heart rate may indicate arousal (an important component of core affect), there have also been efforts to decompose the heart rate signal into a physical and an emotional component[33]. Visser et al.[34] found high heart variability (HRV) was associated with exploratory behaviour, but not freezing behaviour, whereas heart rate was associated with anxious behaviour. Another candidate physiological indicator of positive affect is cortisol. Just as increases in cortisol are likely reflective of acute stress-like states, decreases in cortisol may indicate a transition to more positive, lower arousal, affective states[35]. However studies have found lower levels of faecal cortisol metabolites and plasma cortisol in equines with chronically compromised welfare[36] and increases in cortisol levels can be associated with arousal[37], suggesting that this measure should not be utilised in isolation to measure affective state. Peeters et al.[38] proposed using salivary cortisol as a refinement to serum, given a high correlation between the two when exposed to an adrenocorticotropic (ACTH) challenge. In horses lower respiratory rate has been associated with positive affect for example, Briefer et al.[39] identified a lower respiratory rate in horses listening to affiliative reunion vocalisations. Supporting this Menchetti et al.[40] identified lower respiratory rates in horses with higher scores for calmness. Finally, Waran and Randle[41] proposed that eye temperature may be a suitable metric for assessing affective state. Ramirez Montes De Oca[42] identified changes in the lateralisation of nostril, eye and ear temperatures in dairy calves in rewarded and non-rewarded trials with increasing and decreasing temperatures identified at different anatomical sites, a finding also supported by Kappel[43] in horses. However, evidence in horses is currently limited and requires further validation.

Whilst there are promising candidate measures of positive affective state, research is still at an early stage compared to identification of negative states[44]. To be able to measure affective state it must be reliably induced. One of the challenges of research on affective states is therefore to identify species-specific stimuli that are likely to induce short- or longer-term changes in affective state that are consistent in valence. The ethology of the species combined with prior experience, current environment and internal state will influence whether a particular stimulus induces a positive or negative state. In horses, possibly the best validated induction method is scratching the withers which appears to induce a low arousal positive affect[19,20,45,46]. A further complication is the individual's personality. Strong evidence from the human literature[47–50] and developing evidence from the animal literature[51–56], suggests that personality should be considered within studies of animal affect and this was supported by our prior Delphi consultation[57].

High value food provision is ethologically relevant to equines[58] but has the potential to induce alternative affective responses. For example, it is recognised that high value food provision may induce higher arousal in some equines[59], which may lead to anticipatory behaviours[60] and frustration-related responses[28] rather than a low arousal, positive affective state. In addition, the relative value of perceived high value feedstuff within individuals, according to their personality and prior experiences, is deemed likely to result in a range of affective experiences to induction using this method[61]. The addition of a known affiliative companion has been demonstrated to enhance equine welfare through providing for ethological needs[62–64], potentially inducing positive affect in low arousal. However, understanding the nature of the relationship between the two parties is a crucial factor to ensure that an appropriate selection of social partner is undertaken to induce positive rather than negative affect, and the emotional state of the added conspecific must be considered to avoid instances of negative emotional contagion[65]. Prior research has suggested that positive reinforcement training may enhance short- and longer-term positive affective state[66–68] and increase motivation[69] but could conversely lead to frustration if reinforcement is poorly timed or withheld[28]. Wither scratching is supported within the literature as a potential measure of positive affect induction[19,20,70,71], imitating the relationship building and reassurance seen in allogrooming[58,72]. However, individual preferences have been identified relating to scratching site location[19,45,73] and further individual variation should be considered relating to personality and prior experience when identifying whether an equine is expected to positively respond to human touch[46,74,75].

The aim of this study was to utilise a range of expert suggested methods (wither scratching, high-value food provision, positive reinforcement target training and the addition of an affiliative conspecific)[57] to attempt to induce positive affect in a mixed cohort of equines and to use expert suggested behavioural and physiological measures[57] to evaluate the actual affective state induced in a real-world setting. The study also assessed and

explored the effects of equine personality on the type of affective state induced[76]. Firstly, it was hypothesised that the presentation and subsequent removal of the four induction stimuli would induce differing levels of arousal, and types of valence (positive or negative), within individual horses. We also hypothesised that these affective states (differences in types of valence and levels of arousal) would be evidenced through differences in the behavioural and physiological responses recorded between induction stimuli, and during the phases of the trial (during and after presentation of the stimulus). Thirdly, we hypothesised that the affective states produced via each of the four induction methods may, in part, be affected by individual personality traits. It was also hypothesised that there may be identifiable groupings of measures that may indicate differing levels of arousal and types of valence across the subject cohort.

Results
Reliability of measures

All measures which were recorded by human observers showed good to excellent rater agreement (Respiratory rate intra-rater reliability: ICC, two-way mixed effects, single measures, absolute agreement = 0.995, $p < 0.001$; other behaviour measures inter-rater reliability between 5 raters, ICC, one-way random effects, average measures, absolute agreement = 0.793–0.988, all $p < 0.001$).

Data reduction

From an initial 65 measures, 27 potential measures of affective state were identified to be taken forward (Table 1), based on (1) removal of variables which were highly correlated with other variables; (2) correlation with highly validated QBA measures; (3) support from literature to select between highly correlated measures; (4) ensuring a range of measures across core affect dimensions to distinguish between negatively and positively valenced states and high and low arousal. Measures with low n occurrence were removed. Correlations can be viewed in supplementary information S1 and further statistical and literature-based rationale for measure inclusion can be found in supplementary material S2.

| Measure | Inclusion rationale | | | | Positively correlated variables (≥ 0.5) | Negatively correlated variables (≥ -0.5) |
	Correlation with QBA measures	Literature support	Valence	Arousal		
Ears forward*	0.05	Yes[77]	+	∧/∨	None	None
Ears sideways*	−0.01	Yes[77–79]	+	∨	None	None
Ears asymmetrical*	−0.08	Yes[20,78]	−	∧	Ears gently back	None
Ears gently back*	0.15	Yes[78,79]	+	∨	Ears asymmetrical	None
Ears fully back*	−0.56	Yes[77]	−	∧	Frustration acute conflict	None
Eyes half open*	0.04	Yes[20,80]	+	∨	None	None
Eye sclera*	−0.44	Yes[20,79]	−	∧	None	None
Eye bias both*	−0.20	Yes[81]	+/−	∧/∨	Eye bias right, Eye bias left	None
Eye bias left*	−0.18	Yes[79,82,83]	+/−	∧/∨	Eye bias both	None
Eye bias right*	−0.25	Yes[81]	+/−	∧/∨	Eye bias both	None
Frustration acute conflict	−0.43	Yes[28]	−	∧	Ears fully back	None
Frustration increased locomotion	0.05	Yes[28]	−	∧	None	None
Mean heart rate (BPM)*	0.19	Yes[84]	+/−	∧/∨	None	RR mean
LF power*	−0.02	Yes[85]	+/−	∧	None	HF power
HF power*	−0.03	Yes[85]	+/−	∨	None	LF power
LF:HF ratio*	−0.22	Yes[85]	+/−	∧/∨	None	None
RR mean*	−0.08	Yes[85]	+/−	∧/∨	None	None
pNN50*	0.06	Yes[85]	+/−	∧/∨	RMSSD	None
RMSSD*	0.01	Yes[85]	+/−	∧/∨	SDRR, pNN50	None
SDRR*	−0.01	Yes[85]	+/−	∧/∨	RMSSD	None
Salivary cortisol*	−0.03	Yes[86–91]	+/−	∧/∨	None	None
Respiratory rate*	0.01	Yes[39,40,92–94]	+/−	∧/∨	None	None
QBA Friendly	N/A	Yes[13,95–97]	+	∧/∨	None	None
QBA Relaxed	N/A	Yes[13,95–97]	+	∨	None	None
QBA At ease	N/A	Yes[13,95–97]	+	∨	None	None
QBA Alarmed	N/A	Yes[13,95–97]	−	∧	None	None
QBA Uneasy	N/A	Yes[13,95–97]	−	∧	None	None

Table 1. Potential measures of affective state selected after data reduction procedures, rationale for inclusion and other correlated measures. Indication as to the proposed different dimensions of core affect are represented by positive (+) or negative (−) valence and high (∧) or low (∨) arousal. * Denotes inclusion within later Linear Discriminant Analysis.

Linear discriminant analysis (LDA)

Using wither scratch as the best validated induction method[19,20,46,70,71] to identity potential measures of positive affective state, a linear discriminant analysis was able to distinguish before and during phases (Fig. 1). The variables included in this LDA (denoted by * in Table 1) were taken forward to Principal Component Analysis (PCA), alongside the other measures listed in Table 1.

Principal component analysis (PCA)

In order to consider the hypothesis that there may be identifiable groupings of measures that may indicate differing levels of arousal and types of valence across the subject cohort PCA analysis was applied to all induction methods across all trials (before, during and after) and identified that the first five components accounted for 71% of the variation within the data (PC1: 29%, PC2: 16%, PC3: 13%, PC4: 8%, PC5: 5%). The PCA loading plot was considered in reference to the available literature and using relationships between the selected measures to infer the association with varying valenced states (positive and negative) and levels of arousal (low and high) (Fig. 2). Principal component (PC) 1 (Dim1) appears to be related to the arousal dimension with grouping of measures indicative of high arousal such as 'low-frequency power', 'mean heart rate', 'cortisol', 'ears fully back', 'frustration' and negative 'QBA' measures loading positively on PC1 and groupings of measures indicative of low arousal including 'high-frequency power', 'RR mean', 'ears forward' and positive 'QBA' measures, conversely loading negatively on PC1. PC2 is suggested as a valence dimension with grouping of indicators of potentially negative valence such as 'ears fully back', 'respiratory rate', 'ears asymmetrical', 'eye sclera', negative 'QBA' measures and 'frustration' responses loading negatively and indictors of grouping of measures potentially indicating positive valence such as 'ears forward', 'ears side', positive 'QBA' measures and 'eyes half' loading positively (Fig. 2).

Principal components three and four may represent elements of activation such as seeking (goal-directed engagement with the environment[98]) and conflict (engaging and protective emotions generated by the same experience[99]) potentially representative of positive and negative affective states respectively. PC 3 may be indicative of the activation of seeking type states, potentially in positive valence when aligned with positive PC2 loadings. PC 4 may be indicative of conflicted emotions, potentially in high arousal when aligned with positive PC 1 loadings (Fig. 3). PC 5 appeared unimportant with no trends identified.

Linear modelling of PCA components by stimulus and trial

Considering the hypothesis that the presentation and subsequent removal of the four induction stimuli would induce differing levels of arousal, and types of valence (positive or negative), linear models run on PCA components, did not find consistent changes in components according to induction stimulus and phase. A small, effect of stimulus type was identified within principal component 1 (F = 2.5609, df = 3, p = 0.55; Fig. 4a) but no effect of trial phase was found. No significant effects of trial or phase were identified for principal component 2 (Fig. 4b).

To investigate the hypothesis that affective states (differences in types of valence and levels of arousal) would be evidenced through differences in the behavioural and physiological responses recorded between induction stimuli, and during the phases of the trial (during and after presentation of the stimulus) a mixed model was run, controlling for the effect of individual horses (as random effects), on principal component 1 (arousal) which identified a significant difference, with a large effect size (F value greater than 4) between induction stimuli (Type iii ANOVA with Satterthwaite's method, Stimulus: F = 4.3784, df 3, p = 0.005) with the behavioural and physiological responses to the induction stimulus 'wither scratching' significantly different (with a large effect size, t value greater than 2) when compared with the other three induction stimuli (t = 2.702, df 193.554, p = 0.007). It was identified via evaluation of boxplots (See supplementary material S3) that the frequency rate
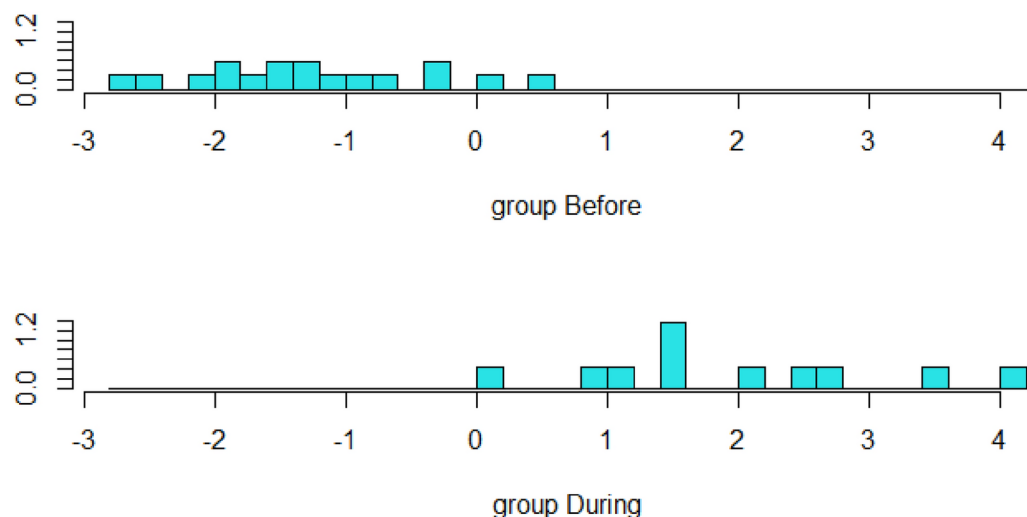

Fig. 1. Separation of variables between the 'before' and 'during' phases of the scratch trial. With correlations of greater than 0.7 identified between variables 'LF power' and 'LF:HF ratio'.
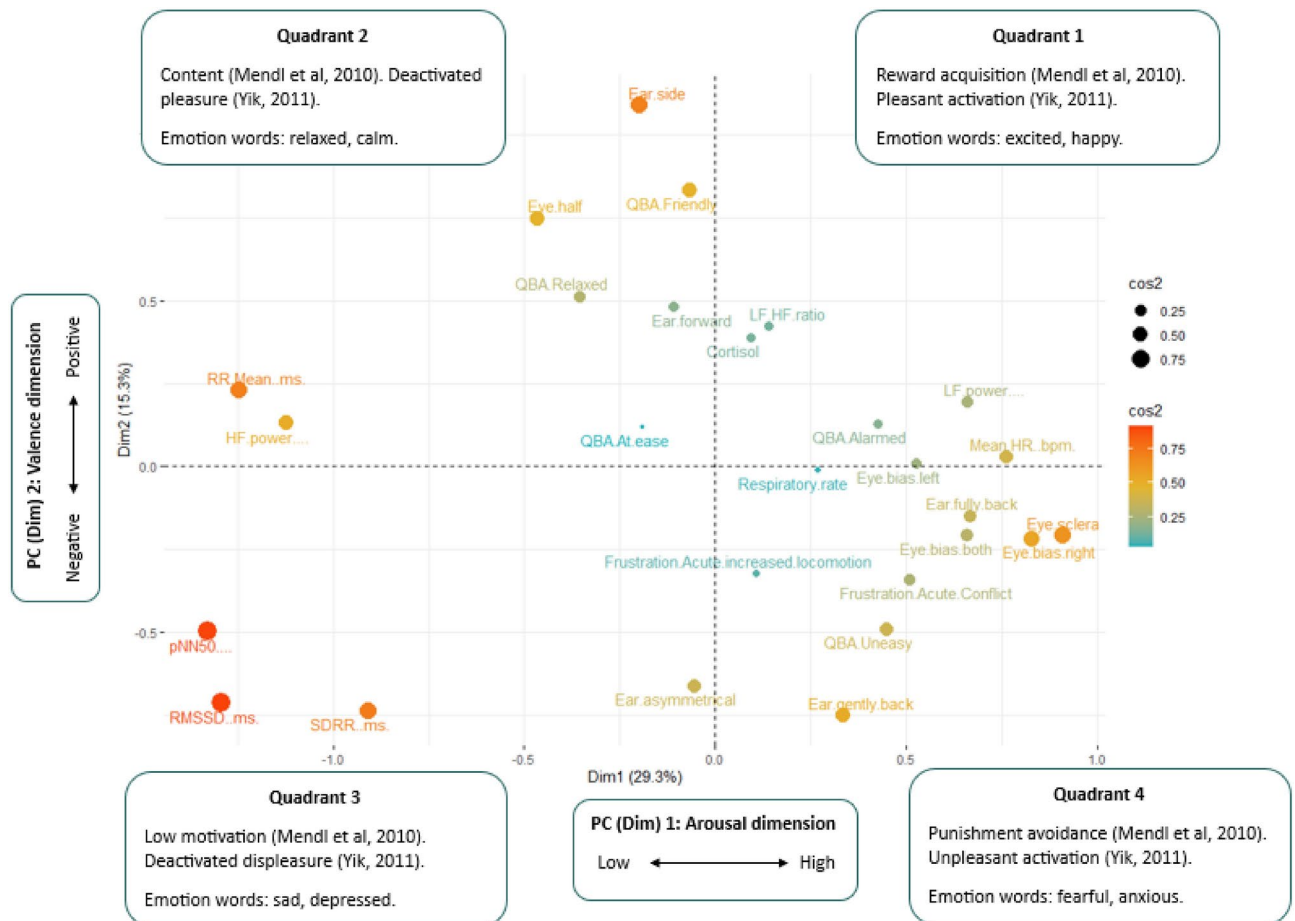
Fig. 2. Proposed model for measures of valence and affect in equines based upon the framework of Mendl et al.[2] following evaluation of measurement variable relationships. Dimension 1 is proposed as the arousal dimension and dimension 2 as the valence dimension. Quadrant 1 is proposed as a higher arousal, positively valenced domain which may be evaluated via assessment of several HR/HRV parameters (Low -Frequency power, LF-HF ratio and Mean heart rate) as well as cortisol levels and potentially one QBA descriptor. Quadrant 2 is proposed as a lower arousal, positively valenced dimension which may be evaluated via assessment of different HRV parameters (RR Mean and High-Frequency power) as well as several QBA descriptors indicative of relaxation and positive valence. Quadrant 3 is proposed as a negatively valenced, lower arousal domain evaluated via assessment of further discrete HRV parameters (pNN50, RMSSD and SDRR) as well as one QBA descriptor. Lastly, quadrant 4 is proposed as a higher arousal, negatively valenced domain evaluated via a number of discrete QBA (Uneasy), behavioural (Frustration, eye bias, eye position and ear position) and physiological (respiratory rate) parameters.

per minute of the behavioural measures 'ears gently back', 'ears asymmetrical', 'eye half' and 'QBA relaxed' were higher during wither scratching when compared with the other three induction stimuli. However, the 'ears side' measure was lower during wither scratching than for the other three induction stimuli. Physiological measure 'salivary cortisol' was lower during wither scratching and the measures 'mean heart rate' and 'pNN50' were lower after wither scratching compared to the other three induction stimuli. 'Ears gently back', 'eye half', 'QBA relaxed' and 'QBA at ease' remained higher after wither scratching compared to the other three induction stimuli. Variable 'RR mean' showed a greater distribution during and after wither scratching compared with the other three induction stimuli.

Within principal component 2 (valence) there was no significant difference in the behavioural and physiological measures between all trial phases and induction stimuli overall (Type iii ANOVA with Satterthwaite's method, Trial: F = 0.5811, df 193.44, p = 0.56; Stimulus: F = 1.3920, df 194.18, p = 0.25). However, there was a significant effect of trial phase on component 2 for 'trial during' (t = 2.525, df 192.86, p = 0.01), 'trial after' (t = 2.130, df 193.42, p = 0.03), and a significant effect of 'stimulus food' (t = 2.898, df 193.81, p = 0.004), 'stimulus scratch' (t = 2.406, df 193.54, p = 0.01) and 'stimulus training' (t = 2.425, df 193.19, p = 0.01); particularly 'trial during stimulus food' (t = − 2.223, df 192.97, p = 0.02), 'trial after stimulus food' (t = − 2.065, df 193.73, p = 0.04) and 'trial during stimulus scratch (t = − 1.995, df 192.86, p = 0.04), indicating that these induction methods and trial phases generated different behavioural and physiological responses across the cohort. Evaluation of boxplots (See supplementary material S3) identified that the behavioural variables 'ears forward', 'gently back' and
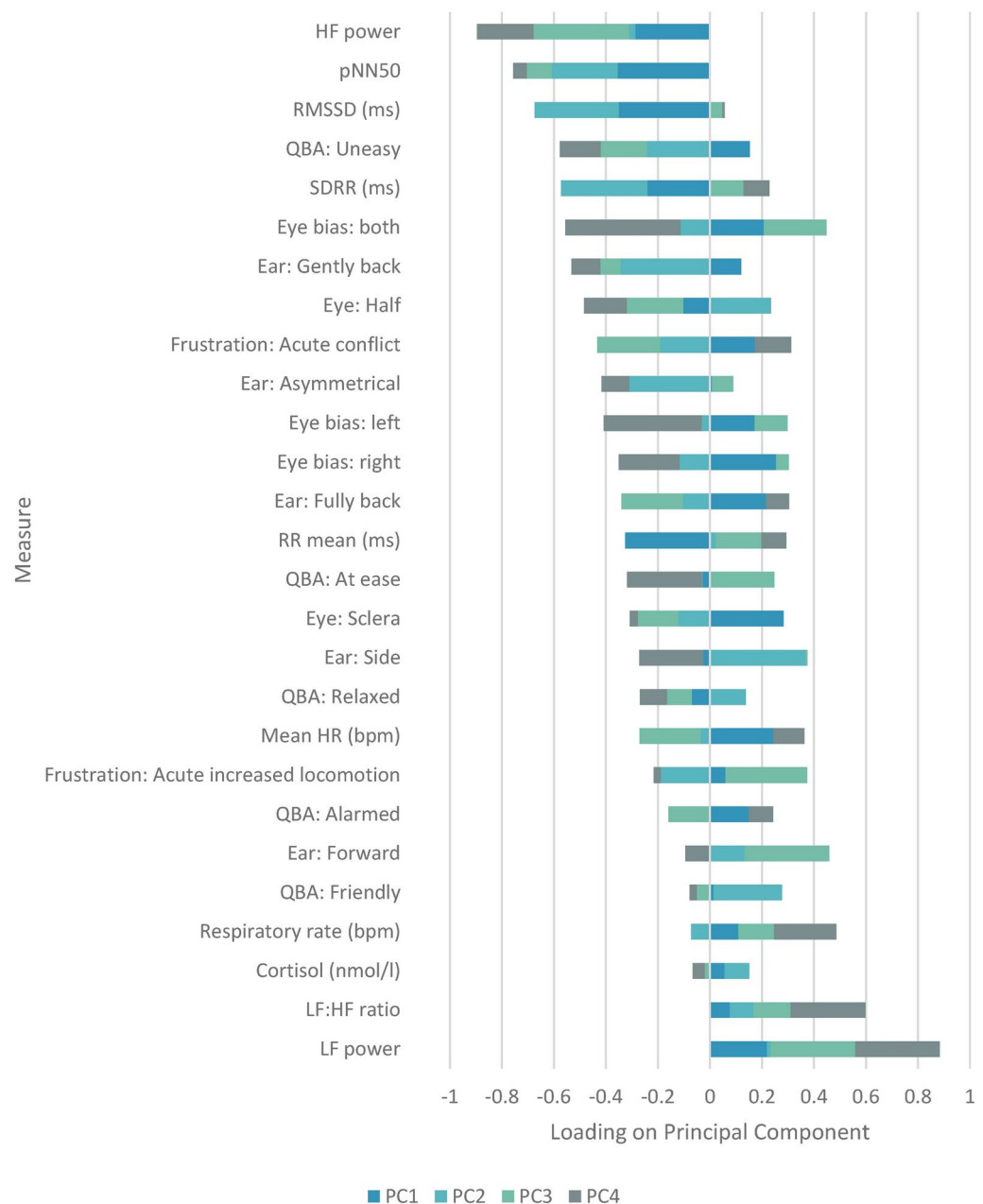
Fig. 3. Loadings of behavioural and physiological measures on principal components 1 to 4.

'asymmetrical' were lower during the food stimulus compared with the other three induction stimuli. 'Eye bias both' and 'left' were lower during the food stimuli than during the other three induction stimuli, as was 'QBA at ease'. After the food stimulus the measures 'ears forward', 'gently back' and 'asymmetrical' were lower than for the other three induction stimuli while 'frustration acute locomotion' and 'HF power' were higher after the food stimulus compared with after the companion, training and wither scratching trials. During the wither scratching trial, it was identified that there were some high outliers within the measure 'eye sclera' alongside a larger 'mean heart rate' and 'RR mean' distribution and low 'RMSSD' distribution. Other differences in measures of note were a high frequency of 'eye bias right', 'frustration acute conflict' and 'QBA uneasy' during the companion stimulus. Interestingly the measure 'frustration acute conflict' increased after all four induction stimuli were removed.

Linear modelling of PCA components by personality group

To investigate the hypothesis that the affective states produced via each of the four induction methods may, in part, be affected by individual personality traits mixed linear models were run to consider the effect of personality on equine responses to trial phase and stimulus which identified some effects of personality traits on components as combined measures of affect (Table 2). These are proposed as increases in arousal displayed by horses within certain personality traits (unfriendly, unresponsive, nervous) to some induction stimuli.
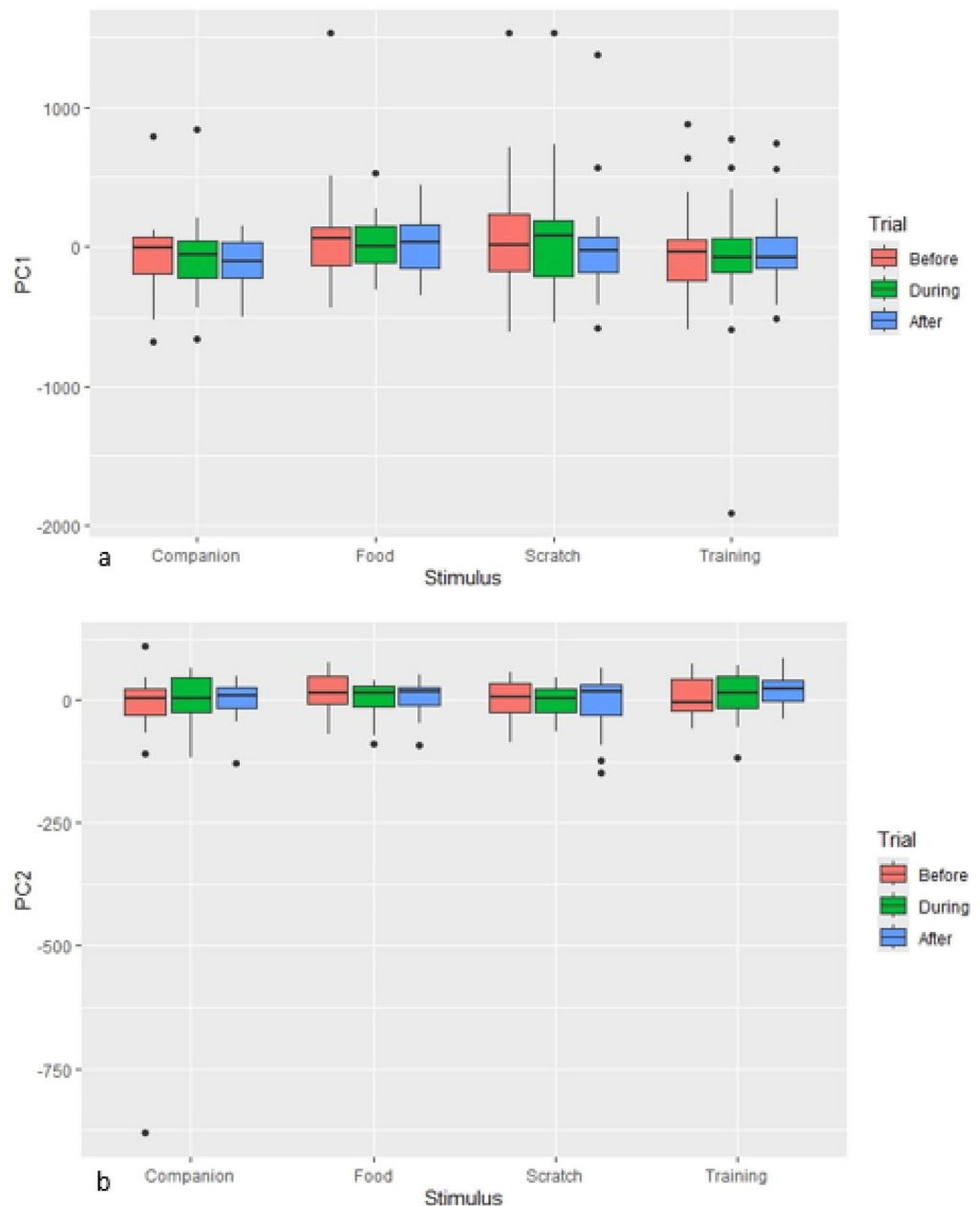
Fig. 4. (**a**) Box and whisker plot displaying loadings on PC1 accounting for the effect of stimulus and trial and (**b**) Box and whisker plot displaying loadings on PC2 accounting for the effect of stimulus and trial.

Additionally, increases in positive valence are suggested within some personality traits (bold, unresponsive) in response to certain induction stimuli.

Discussion

The aim of this study was to induce positive affective state in a cohort of horses to enable identification of potential measures of positive affective state and to differentiate these from potential markers of negative affective state. It was recognised a priori that the induction of positive affect would likely differ between individuals, due in part to their personality traits, as well as the impact of prior life experience, current environment, and learning. The results of this study have identified that the measures utilised to attempt to evaluate affective state in equines appear to be reliable in terms of their inter and intra-rater reliability. Principal component analysis has identified clusters of variables that appear to be informative regarding the affective state of the equines within this study and the variables appear to cluster along the axes of principal components that could be described as valence and arousal components. From the results it appears that some behavioural measures may be useful proxies for their physiological counterparts within the PCA clusters (for example respiratory rate and eye bias left are clustered,

Principal component	Independent variable	Personality trait	Overall model result	Effect variable	Effect result	Suggested affective influence
PC1	**Stimulus**	**Unfriendly**	**F = 5.8926, df = 3, p < 0.001**	**Scratch**	**t = 0.610, df = 199.44, p < 0.001**	**Increased arousal to scratch**
PC1	Stimulus	Affiliative	F = 0.5222, df = 3, p = 0.667	Scratch	t = 2.056, df = 199.68, p < 0.05	Increased arousal to scratch
PC1	Stimulus	Bold	F = 0.6095, df = 3, p = 0.6096	Food	t = 1.993, df = 199.80, p < 0.05	Increased arousal to food
PC1	**Stimulus**	**Unresponsive**	**F = 2.7937, df = 3, p < 0.05**	**Training**	**t = 2.889, df = 200.27, p < 0.05**	**Increased arousal to training**
PC1	**Stimulus**	**Nervous**	**F = 8.7709, df = 3, p < 0.001**	**Food**	**t = 2.765, df = 199.991, p < 0.01**	**Increased arousal to food**
PC1	**Stimulus**	**Nervous**	**F = 8.7709, df = 3, p < 0.001**	**Scratch**	**t = 4.365, df = 200.533, p < 0.001**	**Increased arousal to scratch**
PC2	Stimulus	Exuberant	F = 0.5272, df = 3, p = 0.664	Companion	t = -2.630, df 74.142, p = 0.01	Decreased positive valence to companion
PC2	**Stimulus**	**Bold**	**F = 2.3902, df = 3, p = 0.06**	**Training**	**t = 2.633, df = 199.130, p < 0.01**	Increased positive valence to training
PC2	**Stimulus**	**Bold**	**F = 2.3902, df = 3, p = 0.06**	**Companion**	**t = -2.073, df = 84.244, p < 0.05**	Decreased positive valence to companion
PC2	**Stimulus**	**Unresponsive**	**F = 2.3713, df = 3, p = 0.07**	**Training**	**t = 2.542, df = 201.695, p = 0.01**	Increased positive valence to training
PC2	Stimulus	Nervous	F = 1.7283, df = 3, p = 0.16	Companion	t = -1.999, df = 85.500, p < 0.05	Decreased positive valence to companion
PC2	Stimulus	Nervous	F = 1.7283, df = 3, p = 0.16	Training	t = 2.180, df = 200.734, p < 0.05	Increased positive valence to training
PC2	Trial	Unfriendly	F = 2.4939, df = 2, p = 0.08	After	t = 2.184, df = 204.779, p < 0.05	Increased positive valence after induction stimuli

Table 2. Results of linear models examining the effect of personality traits on responses to stimuli. Results where the overall model was significant are highlighted in bold.

as to some extent are eye half, ear side, QBA relaxed and at ease with RR mean and HF power) and there appears to be an effect of individual personality on equine responses to the four induction stimuli.

Whilst intra-rater agreement was excellent across the respiratory rate measurements these values were not externally tested through inter-rater agreement and as such this reduces the external validity of this measure within this trial. Behavioural data measurements appear extremely reliable, with excellent agreement across all five raters, likely related to the stringent methodology applied for the scoring of behavioural data. Given the range of experience between the raters (from behavioural specialists to raters with no behavioural or equine specialist knowledge), these results appear promising for future use in horses, when behavioural measures are evaluated alongside other parameters.

Prior research has identified LDA analysis as an optimal method for data reduction when taking into account sensitivity, specificity and overall classification accuracy of both physiological and behavioural scientific measures[100]. Here we used wither scratching as the best validated method of inducing affect to identify which measures to apply across all data to make component scores representing aspects of core affect. Wither scratching was selected due to the evidence within the literature[19,20,45,46] and the high rank of this as an induction stimulus by the Delphi experts[57]. The study supported wither scratching as a method for inducing affect since the potential measures of affect were separate between phases of before and during wither scratching when exposed to LDA. A limitation of this approach is that some of the measures could have been specific to wither scratching and therefore less applicable as general indicators of positive affective state.

Examination of the PCA loading plot output identified what appears to be a clear separation between measures that are reported in the literature to account for high and low levels of arousal in PC1 and indicative of positive and negative valence in PC2 (see additional discussion of measures in supplementary material S4). This study was designed around the premise of the core affect 2-dimensional model[2,101,102] with the proposal that different physiological and behavioural measures may be identified which could be proposed as potential measures of varying affective states. The output derived from the PCA plot provides an initial starting point for further exploration of the identified measures of affective state across the four quadrants.

It is postulated that components three and four may be linked to seeking-type motivation and conflicted emotions respectively. Further research could investigate this by measuring plasma dopamine which is positively associated with seeking behaviours in other species[103] and negatively associated with fearfulness in horses[104]. Variables loading positively on PC4 appear to relate to reward acquisition and punishment avoidance[87,90,105] which could be further investigated using reward acquisition or risk-based tasks. This is particularly relevant where frustration may be evident in behavioural outputs[28], as was seen in some equines within this study (with the measure 'acute conflict' increasing after each of the stimuli were removed), which may be related to seeking a different type of interaction (for example with the conspecific) or the removal of reinforcement (for example high value food or the cessation of training). Additionally conflicted emotions, potentially related to con- or hetero-specific interactions during the trial may heighten arousal (potentially with negative valence) thereby dampening any positive effects of the stimulus[106,107]. These are facets that should be considered in future studies.

Interpretation of PCA data is somewhat subjective in terms of attributing meaning to the principal components produced. Relative loadings on components indicates that the variables vary together[108], however, the interpretations made here may not be valid, either due to the over or under estimation of the significance of individual variables or due to inconsistencies that arise within the literature in relation to the potential meaning of variables.

The components did not consistently change in response to the induction methods making it difficult to interpret whether the components were capturing changes in affect, or whether the affective states assumed were not being induced. The four selected induction stimuli had valid rationales for their inclusion however some equines may have found other induction stimuli more intrinsically reinforcing or less negatively valenced or

arousing than those that were selected[109–111]. Individual difference was important in the statistical models, and personality was associated with component scores for some stimuli. This supports our earlier Delphi study[57] in which experts expressed opinions that personality was an important aspect to consider with respect to affective state in horses. Results indicate that consideration should be given to individual personality differences in equines during handling, training and enrichment processes which may alter affective state. Influential personality traits may augment arousal and valence relating to the properties of differing reinforcers meaning that careful examination of personality may enable selection of optimal reinforcement strategies for the individual. For example, our results suggest that equines high in trait 'unresponsive' may show increased arousal in positive valence to target training using positive reinforcement methods, a facet that could be well utilised to improve affective state in these types of equines.

Personality assessment was carried out by owners/carers which may lead to bias. However, these horses and their scores were included in another study[76] where scores were evaluated for both inter and intra-rater reliability, as well as comparisons made between a short and long form of the personality questionnaire; with agreement deemed to range between good and excellent, strengthening the support for the use of these measures within the current study. Additionally, trait ratings such as these have been demonstrated to be substantially more reliable than behaviour coding for assessment of animal personality[112]. It is important to acknowledge the potential for subjective bias within owner scoring of their own animal's personality traits, which may impact upon the reliability of their ratings[113]. Additionally, individual scorers may interpret the personality descriptors differently which may lead to increased variation in scores. Within future research independent personality raters could be included to further test for inter-rater variability and improve the objectivity of personality evaluations, providing that the raters were sufficiently familiar with each study subject. Furthermore cross-validation of personality assessments with behavioural observations during affect induction could strengthen future studies[114].

An important consideration, which is a limitation within this study, is the circular argument that exists where a set of measures were selected to identify positive affective state produced using the selected induction methods where neither measures nor induction methods are all validated. We took several steps to minimise circularity. Firstly, we used a combination of measures which had stronger and weaker evidence that they were related to affective state in horses. By investigating correlations using the PCA between similar (convergent) and dissimilar (divergent) measures we provide a level of concurrent criterion validity (see Williams et al.[115] for how validity terms can be applied to welfare measures). The prior Delphi expert panel[57] could be considered as providing face validity for both measures and induction techniques. By using the best validated induction method, wither scratching, to select measures to take forwards there is a level of concurrent validity. Further studies could look to disentangle arousal and valence components in relation to positive and negative affective states whilst also considering the proposition of neutral affect further within their methodologies. Further investigation of the potential independent, positive linear, symmetric and asymmetric relationships between arousal and valence components, including positivity offset and negativity bias are important facets to consider in this sphere[116]. The identification of physiological and behavioural markers which may be discrete to valence or arousal components could be a useful next step to further understand and assess affective state in equines.

This exploratory study has several limitations which limit the reliability and generalisability of the results reported. The sample size within this study is relatively modest at 20 equines, however, studies like this one have sample sizes within the same range (13–32 horses)[20,43,80,117]. The sample within this study was broadly evenly distributed across sexes, age (young = 0–4 years, mature = 5–15 years and aged = 15 + years) and represented a range of breeds including pure and mixed-breed horses which aids generalisability of the results across equine demographics, but the low sample size prohibits wider generalisation and further analysis of the effects of these demographic differences within the statistical modelling. Further data collection using the same methodological design would add greater power to the study, analysis of a larger dataset may provide instructive evidence around the strength and stability of the observed measurement clusters and retesting of the horses utilised within the current study on at least two further occasions would enable evaluation of the temporal stability of the results reported within this trial. It would be beneficial for researchers in this area to collaborate more widely on a large-scale project to collect a substantial dataset of wide-ranging measurement and induction methods across a much larger cohort of equines to allow for the generation of reliable evidence in this field.

Furthermore, several potential induction stimuli were omitted from the methodological design due to the perceived difficulty in recording the selected behavioural and physiological measures during their implementation. The four selected induction stimuli had valid rationales for their inclusion however it is possible that some equines may have found some of the omitted induction stimuli more intrinsically reinforcing or less negatively valenced or arousing than those that were selected. In particular, some ethologically relevant induction stimuli such as access to graze and browse, turnout and horse-horse interaction at liberty may have provided further useful data in relation to the physiological and behavioural responses to these stimuli and further work could focus on the attempted induction of positive affect using only at liberty, ethologically relevant induction stimuli to reduce the potentially negative effects of confinement and restraint on the achievement of positive affect[109–111].

Conclusion

This study has generated a dataset which provides further advancement of our knowledge around the induction of positive affect and consideration of the meaning and usefulness of variables used to attempt to measure affective state in equines.

The approach was exploratory and provides a novel template for the study of affective state, where there are gaps in knowledge about both methods to induce and the parameters used to measure affective state. In combination the novel aspects of the study were: (i) Selection of measures based on expert opinion and support

from the literature; (ii) Dimension reduction based on the best validated induction methods and comparison with the best validated measures to date; (iii) Inclusion of personality scoring to account for a significant factor likely to lead to variation in equine responses to induction stimuli.

The findings of this work suggest that there is variation within the equine cohort in terms of their physiological and behavioural responses to the induction stimuli that, to some extent, is influenced by personality. The results also suggest variation in the usefulness of the measures utilised for identifying different types of affective state and the dimensional approach taken has allowed for assessment of the broad arousal (high and low) and valence (positive and negative) components of affect without ascribing measures to discrete emotions such as fear for example.

This study is relevant in the current climate where social license to operate has become a significant factor within the equine industry, within which equine welfare is being scrutinised, especially within the competitive sphere. Increased knowledge of how to induce positive affect in horses could improve their welfare and thus prevent the development of pre-pathological and pathological conditions resulting from chronic impaired emotional welfare.

Methods
Ethical approval
The protocol and methods utilised throughout this study were approved by the Animal Welfare Ethics Review Board (AWERB) of Newcastle University (Project ID: 1025), complied with the AWERB guidelines and conformed with ARRIVE guidelines.

Outcomes from the pilot study
A pilot study consisting of 26 horses recruited, with consent, from two university sites was undertaken for refinement of the methodology employed in this study including the consideration of study area set-up, wash-out period between treatments and methods utilised for data processing and analysis. The results of the pilot study identified some interesting trends in relation to changes in both physiological and behavioural parameters in response to different induction stimuli and highlighted the requirement for preference testing of induction methods within the trial to select methods most positively salient to each equine. Furthermore, the pilot study highlighted that a multi-measure approach to measuring affective state would be more reliable than looking at individual measures in isolation.

Trial preparation and subject recruitment
Equine subjects were recruited via individual invitations to their owners. Informed consent was confirmed, and pre-trial demographic and prior history information was gathered for all equines via an online form (See supplementary material S5). Equine personality scoring was undertaken pre-trial, using the Equine Personality Eight-Factor Model (EPEM) full and abridged scoring system[76].

Trials for each equine were conducted across two consecutive days and data collection ran from June to August 2023. Prior to the trial days each subject underwent a food and wither scratching preference testing protocol prior to data collection (See supplementary material, S6).

Subjects
Twenty healthy horses were recruited to the study, from five independent facilities, between March and May 2023 (Table 3).

Trial area set up and habituation
The trial area was set up in an identical manner at each of the five test locations (Fig. 5) with three GoPro Hero 8 cameras (GoPro Inc, USA) placed 3 m to the front of the horse, 1–2 m to the side of the horse and 3 m to the rear of the horse to capture a wide field of view and enable scoring of behavioural and respiratory data.

A thermography camera FLIR T430SC (FLIR LLC, USA) was used to collect thermography data and was time paired with both the GoPro cameras and heart rate monitor. All equines were habituated to the area used for data collection and the equipment used on at least three occasions 24–72 h prior to the trials. Habituation was deemed to have occurred when heart rate did not increase more than 20 beats per minute from baseline during the habituation period[118] and when they quickly (within 1 min) showed postural (relaxed muscle tone, low head carriage) and behavioural (standing and eating hay or resting/dozing) signs of relaxation on entering the area. Horses who did not habituate to the stable area were to be excluded from the trial, however all equines habituated well. All subjects were held on a ~ 10 m loose lead rope for each trial by a known handler to avoid potential stress responses confounding results[119]. The handler was instructed not to interact with the horse during the trial (apart from to place the equine in the correct location for filming) and stood outside the stable and to the right of the closed stable door or through another form of barrier to reduce handler effect[120,121]. A familiar affiliative equine was placed nearby (approximately 10–15 m) during the entire trial to avoid isolation stress[20,122] but not so close as to confound the 'affiliative companion' induction test.

Test procedure
The test protocol was identical for each equine. The subject was led to the research area by a known handler and handled by this handler throughout each trial. The trial process is outlined in full in Fig. 6.

Induction methods
To maximise the likelihood of inducing positive affect in the equine cohort within this study the induction stimuli utilised were those rated most highly by the expert group[57] and supported by the literature. The cross-

Horse number	Age (years)	Breed	Sex	Use	Time in ownership or care	Test location
1	3	Mini Cob	Gelding	Non-ridden	> 6 months < 1 year	1
2	9	Thoroughbred	Gelding	Unaffiliated competition	> 6 years	1
3	5	Irish x Thoroughbred x Arab x Trakehner	Mare	Unaffiliated competition	5–6 years	2
4	16	Warmblood	Gelding	Non-ridden	> 6 months < 1 year	2
5	11	Native crossbreed	Gelding	Unaffiliated competition	1–2 years	2
6	25	Highland	Mare	Hacking	> 6 years	3
7	28	Native crossbreed	Gelding	Non-ridden	> 6 years	3
8	9	Thoroughbred	Gelding	Elite competition	> 6 years	4
9	10	Thoroughbred	Gelding	Elite competition	> 6 years	4
10	4	Thoroughbred	Gelding	Elite competition	1–2 years	4
11	10	Thoroughbred	Gelding	Elite competition	> 6 years	4
12	5	Thoroughbred	Gelding	Elite competition	1–2 years	4
13	26	Cob	Mare	Hacking	> 6 years	5
14	30	Connemara cross	Mare	Non-ridden	> 6 years	5
15	18	Native crossbreed	Mare	Non-ridden	> 6 years	5
16	6	Thoroughbred	Mare	Elite competition	1–2 years	4
17	6	Thoroughbred	Mare	Elite competition	1–2 years	4
18	5	Thoroughbred	Mare	Elite competition	> 6 months < 1 year	4
19	6	Thoroughbred	Gelding	Elite competition	1–2 years	4
20	7	Thoroughbred	Gelding	Elite competition	3–4 years	4

Table 3. Demographic data for the twenty subjects including test location.
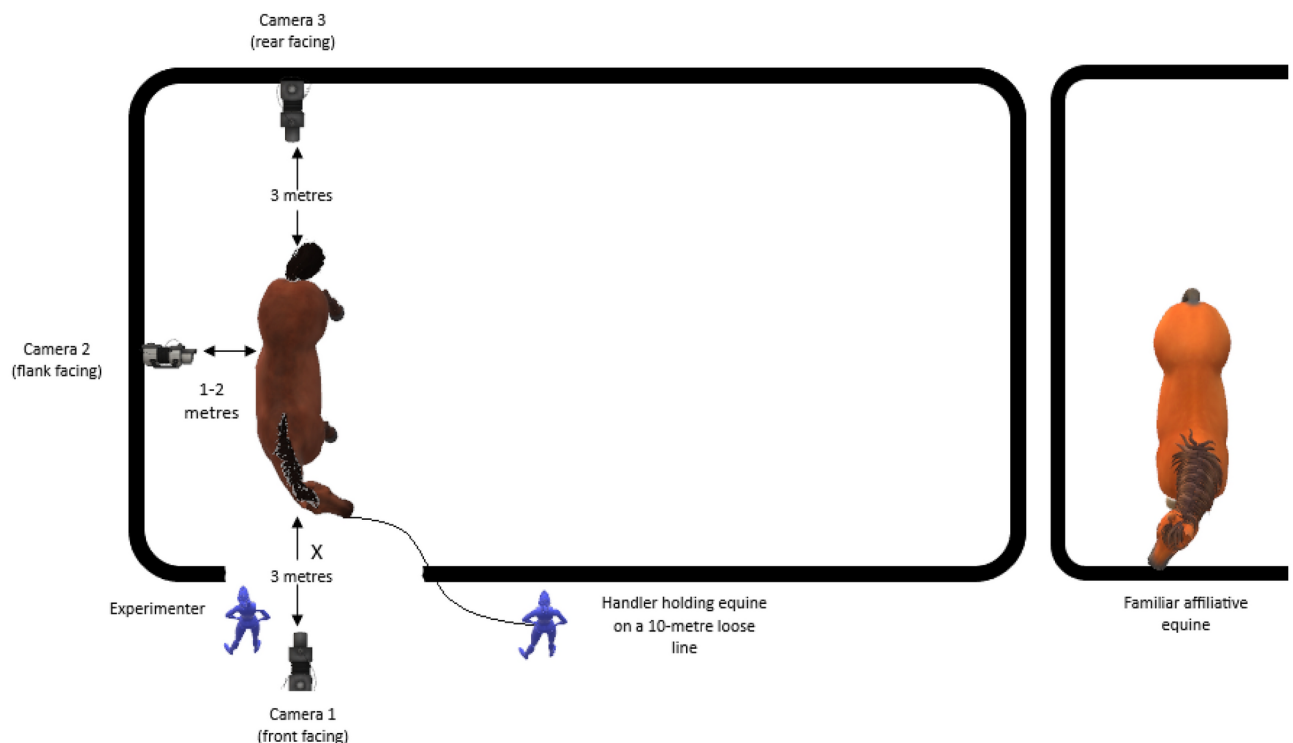

Fig. 5. Trial area set up which was replicated at each location. X marks the position that thermographic images were taken from (1 m in front of the equine).

applicability of induction methods was also considered within the selection process whereby methods that may be highly appetitive to equines but only in a specific or limited circumstance (such as access to a substrate to roll in) were excluded. This process resulted in the selection of the following four induction methods: (i) High value food provision, (ii) The addition of an affiliative conspecific, (iii) Wither scratching and (iv) Application of positive training via target training.

Addition of an affiliative conspecific

Ranked second by the experts and supported within the ethological literature[62–64], this measure was included within the study as a potential induction method for positive affect in low arousal in equines. To utilise the addition of a conspecific to induce positive affect in equines an understanding the nature of the relationship between the two parties is critical.

Within the trial the affiliative conspecific was defined as such following observation of at least five occasions of allogrooming[64] and/or affiliative interactions such as play[64,123] and assessment of proximity to conspecifics, duration of proximity, and the performance of active affiliative behavioural expressions; such as approaches[63,124] with the subject horse in the 4 weeks preceding the trial. The conspecific was brought into view and walked up to the subject horse whereupon the two horses were free to choose to interact or not for the two-minute trial duration. Following this the conspecific was led away out of view.

High value feedstuff

As the highest ranked induction method[57] and one well supported by the literature[59,125] this method was included within the study. This induction stimulus was of particular interest due to its high value, ethologically, to equines[58] but also due to the potential for it to induce alternative affective responses to those aimed for within the study. In addition, the relative value of perceived high value feedstuff within individuals, according to their personality and prior experiences, was deemed likely to result in a range of affective experiences to induction using this method[61], a factor which was identified as potentially useful within later modelling of the measures employed to evaluate affective state achieved. Following preference testing (See supplementary material S6), this trial included the presentation of 1 kg of the preferred substrate in one of the preconditioned buckets[126] which the horse could eat continuously for the two-minute trial period, after which the stimulus was removed.
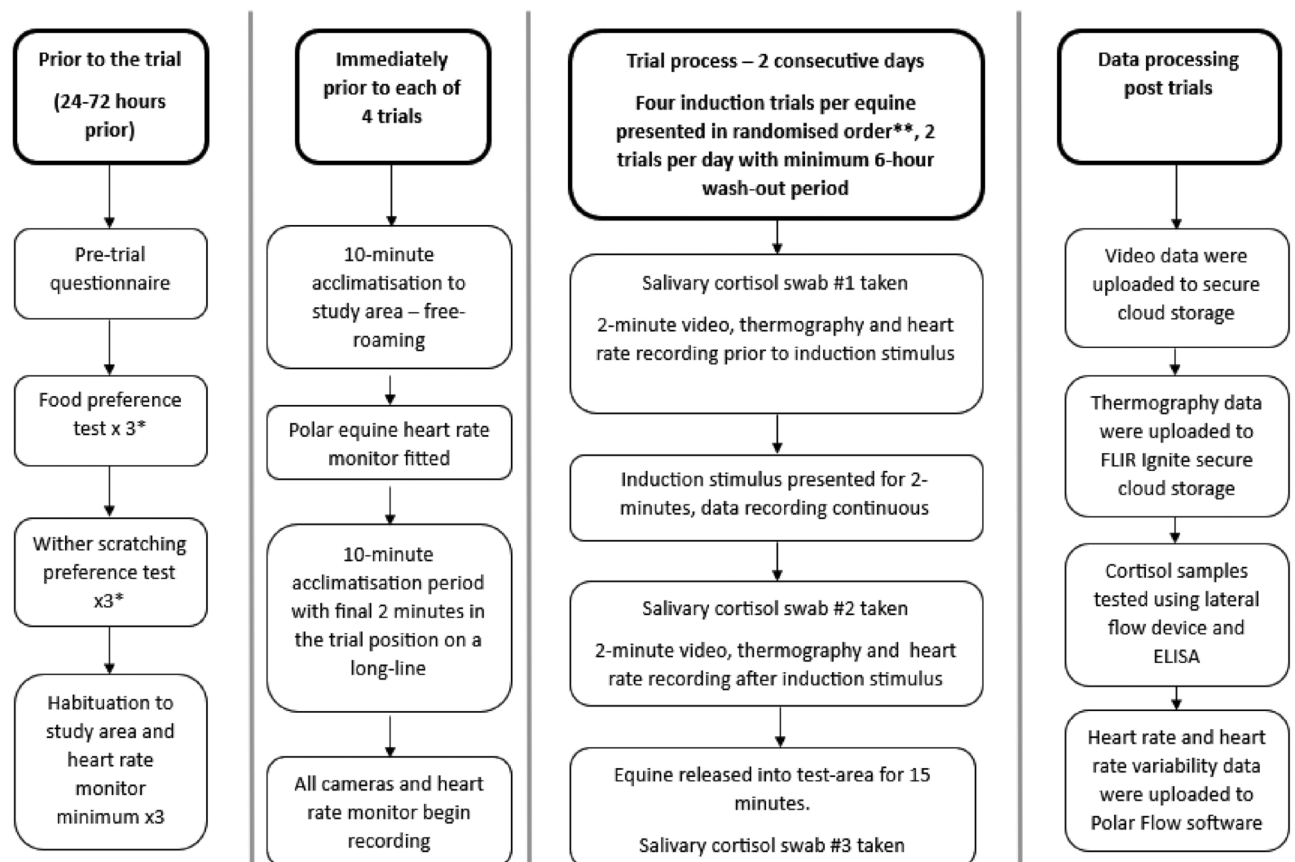

Fig. 6. Schema depicting the pre, during and post-trial processes. * denotes that full methods and results for the preference tests can be found in the supplementary material S6, ** denotes that induction trial stimuli (wither scratching, high value food provision, target training and addition of an affiliative companion) presentation order (randomised) can be found in supplementary material S7.

Wither scratching

Whilst ranked fourth by the experts[57], wither scratching is highly supported within the equine literature as a phenomenon that is broadly experienced as positive by equines. It is thought to contribute to relationship building and comfort[127,128] and has been observed between heterospecifics[58,72] with research identifying links between grooming and/or massaging around the wither area and reduction in heart rate combined with relaxed behavioural responses[19,20,70,71]. Therefore, wither scratching may induce a positive affective state similar to that seen in allogrooming. Prior work has demonstrated a preference for scratching of the wither over other areas such as the neck or face[45] and as such this was selected as the focal scratching area within this study. Within this study wither scratching was trailed with all subjects prior to the trial (See supplementary material S6), and the preferred method and exact site was utilised for each individual subject. The subject received wither scratching, from their known handler, for the entirety of the two-minute trial period and it then abruptly stopped.

Positive reinforcement training

Ranked 9th by the experts[57], this method was deemed to be a novel method which could have good application for inducing positive affect but one that is not widely researched to date. This technique has been previously proposed as a method to optimise equine welfare[66] and although it remains unutilised within equine studies attempting to induce positive affect there is evidence of its efficacy in dogs and wolves[129]. Furthermore, in pigs enhanced attention and cognitive activity has been shown to induce greater positive emotional states[130], a finding that may be transferable to equines and one considered by Sankey et al.[67,68] within their studies looking at reinforcement and positive long-lasting memories in equines, indicating that human attention may be a source of positive emotions for equines. Each equine undertook target training, completed by one of the researchers (LL), for a two-minute duration for this induction trial. The equine was presented with the target stick and upon touching the target with the muzzle, a previously conditioned secondary marker (verbal 'yes' or 'good') was utilised prior to withdrawal of the target and delivery of a low value food reinforcer (chaff) to negate, as much as possible, any crossover in effect between this trial and the high value food trial[59]. Following which the target was presented again to the equine; this process continued for the two-minute trial duration, following which the trainer moved away.

Behavioural and physiological measures

In an earlier phase of this research, we conducted a Delphi consultation[57] to support selection of measures of positive affect for this study. Feedback from the experts indicated consensus agreement that heart rate variability, salivary cortisol, respiratory rate, body language, facial action scoring and qualitative behaviour assessment were reliable methods to assess emotional state in equines. However, many participants commented on the promise but lack of validation of newer methods (eye bias or lateralisation, infra-red thermography of the eye and ear). As such it was decided in this study to utilise a combination of measures which had lesser or greater evidence to validate them as measures of positive affective state and understand how measures covary.

Behavioural measures

Ear and eye position

Ear position (forward, asymmetrical, sideways, gently back, fully back) data were scored for both frequency and duration via continuous recording. Eye positions 'open', 'half open' and 'sclera' were scored for frequency and duration via continuous recording (See supplementary material S8).

Eye bias (lateralisation)

Eye bias was scored continuously for left, both and right eye bias (predominant focus with the left or right eye, or both eyes), with scoring of both duration and frequency undertaken (See supplementary material S8).

Qualitative Behaviour Assessment (QBA)

QBA assessment was undertaken continuously using the full AWIN QBA 13-item descriptor list[95] with both frequency and duration of any behaviours identified (binary presence/absence of the descriptor) recorded for each horse across all trials (See supplementary material S9).

Indicators of frustration

This measure utilised the ten descriptors reported by Pannewitz and Loftus[28] and scored for both frequency and duration of occurrence via continuous sampling[131], (See supplementary material S10).

Physiological measures

Heart rate and heart rate variability (HRV)

From a methodological perspective it is important that when HRV is used to assess emotional state, inter-beat-interval (IBI) data lasting at least five minutes consecutively is measured during stationary conditions to negate the effect of movement behaviour on cardiac function as well as pre-behavioural physiological anticipatory responses and persisting post-behavioural cardiac responses[132]. To satisfy these requirements both heart rate and heart rate variability were recorded continually from the start to the end of each trial phase and for a minimum of 2 min per phase of each trial giving a minimum recording time of 6 min and a maximum of 12 min within this study. All data were recorded using the Polar V800 device and Polar H10 Heart rate sensor band.

Respiratory rate
Respiratory rate was scored via video footage where the rise and fall of the flank was counted as 1 unit. A period of 30 s was scored within the middle section of each trial phase across each induction stimulus for each horse. Scoring was repeated later by the same rater to assess intra-rater reliability.

Salivary cortisol
Within this study each horse acted as its own control for each trial with salivary samples taken before each trial as well as immediately after and 15 min after each trial following a similar test protocol to Kedzierski et al.[133]. Soma equine salivary cortisol test kits (Soma Bioscience, Wallingford, UK) were utilised to take salivary cortisol samples alongside the Soma Cube reader (Soma Bioscience, Wallingford, UK) which analysed lateral flow device (LFD) samples at penside (See supplementary material S11). Subsequently all samples were laboratory analysed via Enzyme Linked Immunosorbent Assay (ELISA)[134] by Soma Bioscience to evaluate the reliability of the LFD analysis.

Ear and eye temperature
Within this study eye and ear temperature were recorded using a Flir T430SC (FLIR Systems AB, Danderyd, Sweden) camera on a moveable tripod set one metre away from the subject. Data were recorded at ~ ten second intervals, and data files uploaded to Flir Ignite online software for analysis. The camera was calibrated to the observation distance, temperature and relative humidity and recalibrated for each subject. Still images of the medial posterior palpebral border of the lower eyelid and lacrimal caruncle of the left and right eyes[135] and the base of each ear[43] were taken at a 90-degree angle in relation to the sagittal plane[43,83] around every 10 s when the horses head was in the optimal position for imaging (example in supplementary material S12).

Personality assessment
Personality assessment was undertaken for each horse as per the method outlined in Loftus et al.[76] All twenty horses included in this final study were rated by their owners or carers on all the individual personality traits and using the full version of the questionnaire and via the abridged questionnaire.

Data processing and statistical analysis
Data processing
Cortisol data. Cronbach's Alpha reliability analysis was undertaken to compare the LFD and Elisa results which showed moderate agreement ($r^2 = 0.4$, $p < 0.001$). Therefore to ensure accuracy the Elisa results were taken forward for analysis.

Heart rate data. Heart rate data were extracted from Polar Flow individually per horse, per stimulus. The data were then uploaded to Kubios version 3.5.0 (Kubios OY, Finland)[136]. Data were treated using the method outlined by Scopa et al.[85], initially the R peaks of the QRS complexes were determined using Kubios' inbuilt QRS detector and the Pan-Tompkins algorithm method. Data were passed through a bandpass filter (frequency band of 0.05–32 Hz and window width of 150 ms) to smooth close-by peaks. Medium correction was utilised to avoid over-correcting and missing normal inter-beat intervals. Three spectral bands were identified that have been specifically recommended for equine HRV analysis[85]; Very Low-frequency (VLF) was set to 0–0.01 Hz, Low-frequency (LF) was set to 0.01–0.07 Hz and High-frequency was set to 0.07–0.6 Hz.

In the time domain, Mean RR, SDNN, Mean HR, SDHR, Minimum HR, Maximum HR, RMSSD and pNN50 data were extracted. In the frequency domain, AR values were extracted as these are most applicable for data collected over shorter timeframes. These included Peak (Hz), Power (ms^2), Power (%) and Power (n.u.) for the VLF, LF and HF domains; LF-HF ratio was also computed. In the non-linear domain, SD1 and SD2 were computed.

Respiratory rate data. Respiratory rate was calculated from video data extracted from a side-facing GoPro camera which was mounted at a height and position to capture the rise and fall of the flanks. Each rise and fall were counted as one respiratory unit. Respiratory rate was counted using a four-digit clicker counter (Sdarming, Dongguan Xunze Trading Company, China) for 30 s in the middle section of each phase of each trial and the number was doubled to give a beat per minute value. Intra-observer reliability was tested through rescoring of 20% of the counts 2 weeks following the initial count.

Thermography data. Still images from each phase of each stimulus presentation were analysed using Flir Ignite software. Initially, images were inspected for clarity, focus and availability of measurement of each of the four data points. Eye temperature was taken from the medial canthus of the eye as suggested by Kim and Cho[135].

Behavioural data. All behavioural measures were scored by the principal researcher (LL) using ethograms constructed prior to scoring (See supplementary material S8-10) and 20% of videos were tested for inter-rater reliability.

Behavioural data were scored from video using CowLog (v3.0.2) software[137,138] and followed a predetermined method (See supplementary material S13).

Once scored the behavioural data were placed into Microsoft Excel v16 (Microsoft Corporation, 2016) with pre-set formulas which calculated total duration and frequency of each type of scored behaviour.

Inter-observer reliability was tested utilising intraclass correlation coefficient (ICC) analysis on data from 5 trained additional raters who between them re-scored 20% of the videos (at least one per horse and three per stimulus). Two raters were specialists in equine behaviour, two raters were equine specialists but not specialists in behaviour and the final rater was neither an equine nor a behaviour specialist. All raters received the same training, scoring software, ethograms and scoring instructions and all scored the videos independently of each other. Random allocation of videos to the raters was determined using online randomisation software (www.random.org).
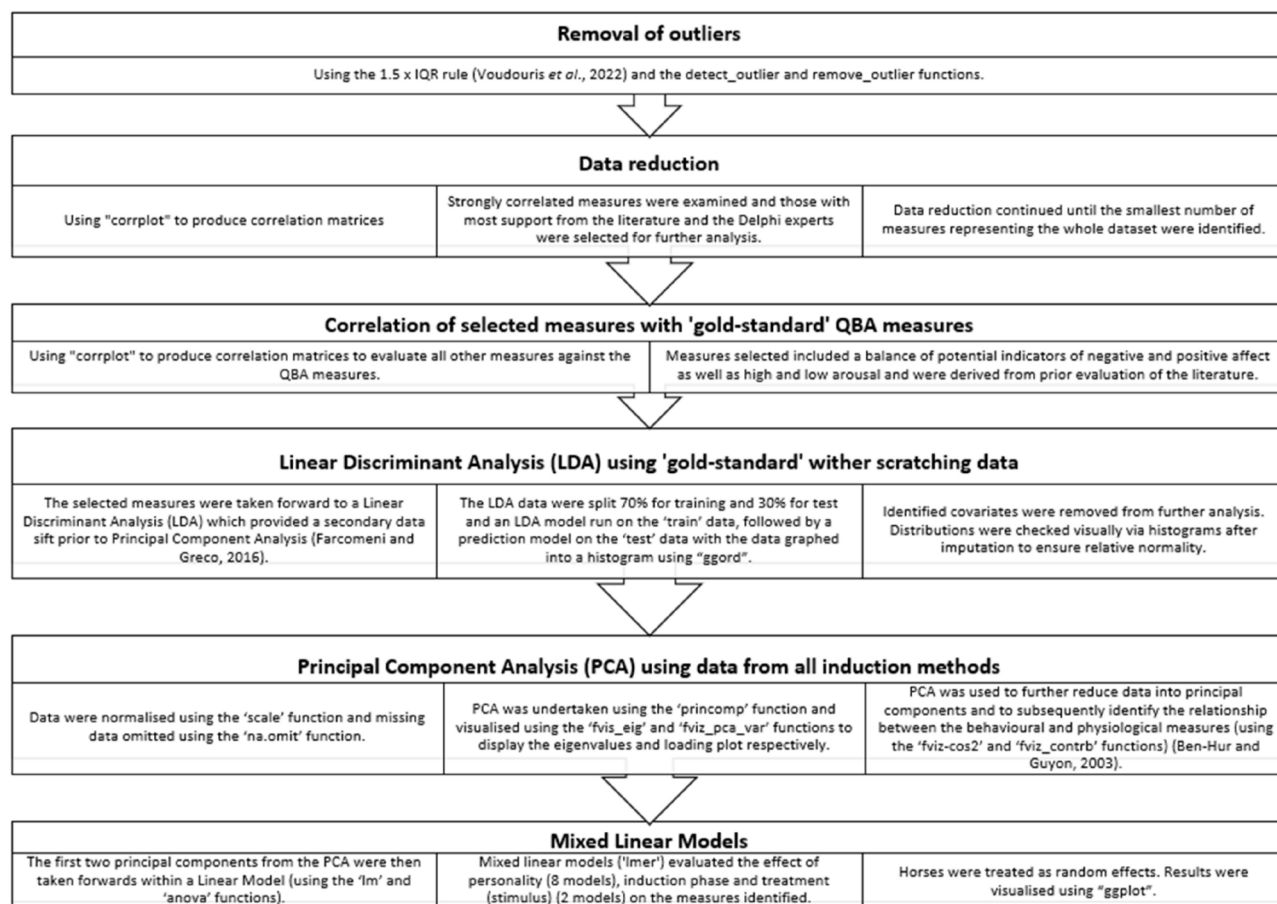
Fig. 7. Flow of statistical analysis undertaken.

Statistical analysis

An exploratory approach was undertaken throughout the statistical analysis to attempt to identify behavioural and physiological markers of affective state (both positive and negative) and to consider the effect of the treatment (stimulus), personality and the random effect of the horse within analyses. Four a priori assumptions were made (i) There would be covariation of measures[139–142]; (ii) Wither scratching induces positive affect[19,20,70], (iii) Personality would influence behavioural and physiological responses to different induction stimuli[143,144] and iv) QBA measures are valid indicators due to their prior validation[13,95,145]. These assumptions were made following review of the literature which has previously identified covariation between some behavioural and physiological measures of equine affective state[19,20] and identified wither scratching as a method which is likely to induce relaxation and positive emotions in horses[20,45,46,146]. Evidence from the literature also identified the likely influence of personality traits on equine behavioural responses[147–149].

All analyses were undertaken using R Studio 12.1[150] utilising the packages "car", "corr", "corrplot", "devtools", "FactoMineR", "factoextra", "fansi", "ggcorrplot", "ggord", "KlaR", "lme4", "MASS" and "Psych". Initially, outliers were removed from the dataset using the $1.5 \times$ IQR rule[151] and the detect_outlier and remove_outlier functions; following which correlation matrices were computed using "corrplot" to identify measures which appeared to covary to allow for later data reduction[152] (See supplementary material S1). Variables with a high number of missing values (> 10 per variable) were excluded from analysis. This included several of the QBA and frustration measures and where variables contained low levels of missing values (≤ 10 per variable) the median value for the variable was computed and imputed in place of the missing values (this occurred mainly for thermography and HRV data). Distributions were checked visually via histograms after imputation to ensure relative normality. The statistical methodology is described in full in Fig. 7.

The QBA data frame was scaled and normalised. The well validated QBA measures[13,95,97] were evaluated for correlations (using corrplot) with the other measures (ear data, eye data, frustration data and heart rate data) (See supplementary material S1). A balance of measures with both negative and positive correlations to the QBA variables were selected to be taken forward using evidence for their value from the literature. Following initial review of the relationships between variables it was noted that there did not seem to be sufficient variables included within the PCA to meaningfully distinguish between potential indicators of negative and positive affective states. Therefore, further variables were included within the PCA analysis (See supplementary material S2). Cortisol and respiratory rate data were also added as physiological variables of interest as highlighted by the literature and as there appeared to be some differences between treatments and trial phase within the cortisol

and the respiratory data collected within this study (See supplementary material S14). As an additional data sift an LDA was run using only the wither scratch data with the rationale that wither scratching was the most highly validated induction method for calm, positive affect according to evidence from the literature[19,20,45,46]. The LDA was only run on phases 'before' and 'during' to avoid the impact of any carry over effect during phase 'after' which may have skewed the data[153]. Initially, linear models were run using principal components 1 and 2 to identify any effect of trial phase (before, during and after) and stimulus (food, wither scratching, companion and training) on the variables within each component. Following this, mixed linear models were run to identify any effect horse as a random effect (2 models, one for PC1 and one for PC2) on the treatments evaluated (trial phase and stimulus). Subsequently, eight mixed models were run for PC1 (one for each of the eight personality traits, which were considered fixed effects), and the same process was repeated for PC2 to identify any effect of personality trait on the variables within the principal components. Following the linear models on the principal components, data on relevant behaviours were visualised by stimulus and trial phase to gain more detail of which specific behaviour changed in response to the stimuli (See supplementary material S3).

Data availability

The dataset analysed during the current study is available from the corresponding author on reasonable request.

References

1. Mendl, M. & Paul, E. S. In *Mental Health and Well-Being in Animals* (ed. McMillan, F. D.) 328–344 (CABI, 2020).
2. Mendl, M., Burman, O. H. P. & Paul, E. S. An integrative and functional framework for the study of animal emotion and mood. *Proc. R. Soc.* **277**, 2895–2904. https://doi.org/10.1098/rspb.2010.0303 (2010).
3. Boissy, A. & Lee, C. How assessing relationships between emotions and cognition can improve farm animal welfare. *Rev. Sci. Tech.* **33**, 103–110. https://doi.org/10.20506/rst.33.1.2260 (2014).
4. Blumenthal, A. L. *A Reappraisal of Wilhelm Wundt* (American Psychological Association, 2002).
5. Reimert, I., Webb, L. E., van Marwijk, M. A. & Bolhuis, J. E. Review: Towards an integrated concept of animal welfare. *Animal* **17**, 100838. https://doi.org/10.1016/j.animal.2023.100838 (2023).
6. Douglas, J., Owers, R. & Campbell, M. H. Social licence to operate: What can equestrian sports learn from other industries?. *Animals* **12**, 1987. https://doi.org/10.3390/ani12151987 (2022).
7. Heleski, C. R. Social license to operate—why public perception matters for horse sport—some personal reflections. *J. Equine Vet.* **124**, 104266. https://doi.org/10.1016/j.jevs.2023.104266 (2023).
8. Paul, E. S., Sher, S., Tamietto, M., Winkielman, P. & Mendl, M. T. Towards a comparative science of emotion: Affect and consciousness in humans and animals. *Neurosci. Biobehav. Rev.* **108**, 749–770. https://doi.org/10.1016/j.neubiorev.2019.11.014 (2020).
9. Boissy, A. et al. Emotions and cognition: A new approach to animal welfare. *Anim. Welf.* **16**, 37–43. https://doi.org/10.1017/S0962728600031717 (2007).
10. Panksepp, J. The basic emotional circuits of mammalian brains: Do animals have affective lives?. *Neurosci. Biobehav. Rev.* **35**, 1791–1804. https://doi.org/10.1016/j.neubiorev.2011.08.003 (2011).
11. Harding, E. J., Paul, E. S. & Mendl, M. Cognitive bias and affective state. *Nature* **427**, 312–312. https://doi.org/10.1038/427312a (2004).
12. Mendl, M., Burman, O. H. P., Parker, R. M. A. & Paul, E. S. Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. *Appl. Anim. Behav. Sci.* **118**, 161–181. https://doi.org/10.1016/j.applanim.2009.02.023 (2009).
13. Dalla Costa, E. et al. Welfare assessment of horses: The AWIN approach. *Anim. Welf.* **25**, 481–488. https://doi.org/10.7120/09627286.25.4.481 (2016).
14. Temple, D., Manteca, X., Velarde, A. & Dalmau, A. Assessment of animal welfare through behavioural parameters in Iberian pigs in intensive and extensive conditions. *Appl. Anim. Behav. Sci.* **131**, 29–39. https://doi.org/10.1016/j.applanim.2011.01.013 (2011).
15. Andreasen, S. N., Wemelsfelder, F., Sandøe, P. & Forkman, B. The correlation of Qualitative Behavior Assessments with Welfare Quality® protocol outcomes in on-farm welfare assessment of dairy cattle. *Appl. Anim. Behav. Sci.* **143**, 9–17. https://doi.org/10.1016/j.applanim.2012.11.013 (2013).
16. Hall, C. & Heleski, C. The role of the ethogram in equitation science. *Appl. Anim. Behav. Sci.* **190**, 102–110. https://doi.org/10.1016/j.applanim.2017.02.013 (2017).
17. Draaisma, R. *Language Signs and calming Signals of Horses: Recognition and Application* (CRC Press, 2017).
18. Wathan, J., Burrows, A., Waller, B. M. & McComb, K. EquiFACS: The equine facial action coding system. *PLoS ONE* **10**, e0131738. https://doi.org/10.1371/journal.pone.0131738 (2015).
19. Lansade, L., Lemarchand, J., Reigner, F., Arnould, C. & Bertin, A. Automatic brushes induce positive emotions and foster positive social interactions in group-housed horses. *Appl. Anim. Behav. Sci.* **246**, 105538. https://doi.org/10.1016/j.applanim.2021.105538 (2022).
20. Lansade, L. et al. Facial expression and oxytocin as possible markers of positive emotions in horses. *Sci. Rep.* **8**, 1–11. https://doi.org/10.1038/s41598-018-32993-z (2018).
21. Trösch, M. et al. Horses feel emotions when they watch positive and negative horse–human interactions in a video and transpose what they saw to real life. *Anim. Cogn.* **23**, 643–653. https://doi.org/10.1007/s10071-020-01369-0 (2020).
22. Rogers, L. J., Vallortigara, G. & Andrew, R. J. *Divided Brains: The Biology and Behaviour of Brain Asymmetries* (Cambridge University Press, 2013).
23. Baragli, P., Scopa, C., Felici, M. & Reddon, A. R. Horses show individual level lateralisation when inspecting an unfamiliar and unexpected stimulus. *PLoS ONE* **16**, e0255688. https://doi.org/10.1371/journal.pone.0255688 (2021).
24. Felici, M., Reddon, A. R., Maglieri, V., Lanatà, A. & Baragli, P. Heart and brain: Change in cardiac entropy is related to lateralised visual inspection in horses. *PLoS ONE* **18**, e0289753. https://doi.org/10.1371/journal.pone.0289753 (2023).
25. Amsel, A. *Frustration Theory: An Analysis of Dispositional Learning and Memory* (Cambridge University Press, 1992).
26. Dollard, J. et al. *Frustration and Aggression* Vol. 4 (Routledge, 2013).
27. Hintze, S., Murphy, E., Bachmann, I., Wemelsfelder, F. & Würbel, H. Qualitative Behaviour Assessment of horses exposed to short-term emotional treatments. *Appl. Anim. Behav. Sci.* **196**, 44–51. https://doi.org/10.1016/j.applanim.2017.06.012 (2017).
28. Pannewitz, L. & Loftus, L. Frustration in horses: Investigating expert opinion on behavioural indicators and causes using a Delphi consultation. *Appl. Anim. Behav. Sci.* **258**, 105818. https://doi.org/10.1016/j.applanim.2022.105818 (2023).

29. Munsters, C. C. B. M., Visser, E. K., van den Broek, J. & van Oldruitenborgh-Oosterbaan, M. M. S. The influence of challenging objects and horse-rider matching on heart rate, heart rate variability and behavioural score in riding horses. *Vet. J.* **192**, 75–80. https://doi.org/10.1016/j.tvjl.2011.04.011 (2012).

30. von Lewinski, M. et al. Cortisol release, heart rate and heart rate variability in the horse and its rider: Different responses to training and performance. *Vet. J.* **197**, 229–232. https://doi.org/10.1016/j.tvjl.2012.12.025 (2013).

31. Villas-Boas, J. D. et al. Behavioural, endocrine and cardiac autonomic responses to a model of startle in horses. *Appl. Anim. Behav. Sci.* **174**, 76–82. https://doi.org/10.1016/j.applanim.2015.10.005 (2016).

32. Schmidt, A., Aurich, J., Möstl, E., Müller, J. & Aurich, C. Changes in cortisol release and heart rate and heart rate variability during the initial training of 3-year-old sport horses. *Horm. Behav.* **58**, 628–636. https://doi.org/10.1016/j.yhbeh.2010.06.011 (2010).

33. Jansen, F. et al. Online detection of an emotional response of a horse during physical activity. *Vet. J.* **181**, 38–42. https://doi.org/10.1016/j.tvjl.2009.03.017 (2009).

34. Visser, E. et al. Responses of horses in behavioural tests correlate with temperament assessed by riders. *Equine Vet. J.* **35**, 176–183. https://doi.org/10.2746/042516403776114108 (2003).

35. Buchanan, T. W., Al'Absi, M. & Lovallo, W. R. Cortisol fluctuates with increases and decreases in negative affect. *Psychoneuroendocrinology* **24**, 227–241. https://doi.org/10.1016/S0306-4530(98)00078-X (1999).

36. Pawluski, J. et al. Low plasma cortisol and fecal cortisol metabolite measures as indicators of compromised welfare in domestic horses (*Equus caballus*). *PLoS ONE* **12**, e0182257. https://doi.org/10.1371/journal.pone.0182257 (2017).

37. Ralph, C. R. & Tilbrook, A. J. Invited review: The usefulness of measuring glucocorticoids for assessing animal welfare. *J. Anim. Sci.* **94**, 457–470. https://doi.org/10.2527/jas.2015-9645 (2016).

38. Peeters, M., Sulon, J., Beckers, J. F., Ledoux, D. & Vandenheede, M. Comparison between blood serum and salivary cortisol concentrations in horses using an adrenocorticotropic hormone challenge. *Equine Vet. J.* **43**, 487–493. https://doi.org/10.1111/j.2042-3306.2010.00294.x (2011).

39. Briefer, E. F. et al. Perception of emotional valence in horse whinnies. *Front. Zool.* **14**, 1–12. https://doi.org/10.1186/s12983-017-0193-1 (2017).

40. Menchetti, L., Dalla Costa, E., Minero, M. & Padalino, B. Development and validation of a test for the classification of horses as broken or unbroken. *Animals* **11**, 2303. https://doi.org/10.3390/ani11082303 (2021).

41. Waran, N. & Randle, H. What we can measure, we can manage: The importance of using robust welfare indicators in Equitation Science. *Appl. Anim. Behav. Sci.* **190**, 74–81. https://doi.org/10.1016/j.applanim.2017.02.016 (2017).

42. Ramirez Montes De Oca, M. A. Temperature asymmetries in facial areas as indicators of affective state in dairy calves and horses. PhD thesis, University of Bristol (2021).

43. Kappel, S. Towards the validation of novel cognitive and physiological indicators of affective state in horses (*Equus caballus*). PhD thesis, University of Plymouth (2023).

44. Hall, C., Randle, H., Pearson, G., Preshaw, L. & Waran, N. Assessing equine emotional state. *Appl. Anim. Behav. Sci.* **205**, 183–193. https://doi.org/10.1016/j.applanim.2018.03.006 (2018).

45. Thorbergson, Z. W., Nielsen, S. G., Beaulieu, R. J. & Doyle, R. E. Physiological and behavioral responses of horses to wither scratching and patting the neck when under saddle. *J. Appl. Anim. Welf. Sci.* **19**, 245–259. https://doi.org/10.1080/10888705.2015.1130630 (2016).

46. Watson, J. C. & McDonnell, S. M. Effects of three non-confrontational handling techniques on the behavior of horses during a simulated mildly aversive veterinary procedure. *Appl. Anim. Behav. Sci.* **203**, 19–23. https://doi.org/10.1016/j.applanim.2018.02.007 (2018).

47. Peterson, C. B. et al. Psychometric properties of the eating disorder examination-questionnaire: Factor structure and internal consistency. *Int. J. Eat. Disord.* **40**, 386–389. https://doi.org/10.1002/eat.20373 (2007).

48. Zuckerman, M. In *Advanced Personality* (eds Barone, D. F. et al.) 123–154 (Springer, 1998).

49. Akiskal, H. S. In *Psychosocial Aspects of Depression* (eds Becker, J. & Kleinman, A.) 215–236 (Routledge, 1990).

50. Gunnar Götestam, K., Svebak, S. & Naper Jensen, E. The role of personality, mood, subjective health, and stress in depressive symptoms among high school students. *Eur. J. Psychiatry* **22**, 121–129. https://doi.org/10.4321/S0213-61632008000300001 (2008).

51. Mills, D. S. Personality and individual differences in the horse, their significance, use and measurement. *Equine Vet. J.* **30**, 10–13. https://doi.org/10.1111/j.2042-3306.1998.tb05137.x (1998).

52. Gosling, S. D. Personality in non-human animals. *Soc. Pers. Psychol. Compass* **2**, 985–1001. https://doi.org/10.1111/j.1751-9004.2008.00087.x (2008).

53. Hausberger, M., Muller, C. & Lunel, C. Does work affect personality? A study in horses. *PLoS ONE* **6**, e14659. https://doi.org/10.1371/journal.pone.0014659 (2011).

54. Sih, A. et al. Animal personality and state–behaviour feedbacks: A review and guide for empiricists. *Trends Ecol. Evol.* **30**, 50–60. https://doi.org/10.1016/j.tree.2014.11.004 (2015).

55. Gartner, M. C., Powell, D. M. & Weiss, A. Comparison of subjective well-being and personality assessments in the clouded leopard (*Neofelis nebulosa*), Snow Leopard (*Panthera uncia*), and African Lion (*Panthera leo*). *J. Appl. Anim. Welf. Sci.* **19**, 294–302. https://doi.org/10.1080/10888705.2016.1141057 (2016).

56. Robinson, L. M., Waran, N. K., Handel, I. & Leach, M. C. Happiness, welfare, and personality in rhesus macaques (*Macaca mulatta*). *Appl. Anim. Behav. Sci.* **236**, 105268. https://doi.org/10.1016/j.applanim.2021.105268 (2021).

57. Loftus, L., Asher, L. & Leach, M. C. Inducing and measuring positive affective state in equines: A Delphi consultation (In prep).

58. McGreevy, P. *Equine Behavior: A Guide for Veterinarians and Equine Scientist* 2nd edn. (Elsevier, 2012).

59. Kieson, E., Felix, C., Webb, S. & Abramson, C. I. The effects of a choice test between food rewards and human interaction in a herd of domestic horses of varying breeds and experiences. *Appl. Anim. Behav. Sci.* **231**, 105075. https://doi.org/10.1016/j.applanim.2020.105075 (2020).

60. Peters, S. M., Bleijenberg, E. H., van Dierendonck, M. C., van der Harst, J. E. & Spruijt, B. M. Characterization of anticipatory behaviour in domesticated horses (*Equus caballus*). *Appl. Anim. Behav. Sci.* **138**, 60–69. https://doi.org/10.1016/j.applanim.2012.01.018 (2012).

61. Elia, J. B., Erb, H. N. & Houpt, K. A. Motivation for hay: Effects of a pelleted diet on behavior and physiology of horses. *Physiol. Behav.* **101**, 623–627. https://doi.org/10.1016/j.physbeh.2010.09.010 (2010).

62. Kimura, R. Mutual grooming and preferred associate relationships in a band of free-ranging horses. *Appl. Anim. Behav. Sci.* **59**, 265–276. https://doi.org/10.1016/S0168-1591(97)00129-9 (1998).

63. Heitor, F., do Mar Oom, M. & Vicente, L. Social relationships in a herd of Sorraia horses. Part II Factors affecting affiliative relationships and sexual behaviours. *Behav. Process.* **73**, 231–239. https://doi.org/10.1016/j.beproc.2006.05.005 (2006).

64. Ransom, J. I. & Cade, B. S. Quantifying equid behavior—a research ethogram for free-roaming feral horses (US Geological Survey Techniques and Methods, 2009).

65. Spinka, M. Social dimension of emotions and its implication for animal welfare. *Appl. Anim. Behav. Sci.* **138**, 170–181. https://doi.org/10.1016/j.applanim.2012.02.005 (2012).

66. Hausberger, M. et al. Mutual interactions between cognition and welfare: The horse as an animal model. *Neurosci. Biobehav. Rev.* **107**, 540–559. https://doi.org/10.1016/j.neubiorev.2019.08.022 (2019).

67. Sankey, C., Richard-Yris, M.-A., Leroy, H., Henry, S. & Hausberger, M. Positive interactions lead to lasting positive memories in horses, *Equus caballus*. *Anim. Behav.* **79**, 869–875. https://doi.org/10.1016/j.anbehav.2009.12.037 (2010).

68. Sankey, C. et al. Reinforcement as a mediator of the perception of humans by horses (*Equus caballus*). *Anim. Cogn.* **13**, 753–764. https://doi.org/10.1007/s10071-010-0326-9 (2010).
69. Innes, L. & McBride, S. D. Negative versus positive reinforcement: An evaluation of training strategies for rehabilitated horses. *Appl. Anim. Behav. Sci.* **112**, 357–368 (2008).
70. Feh, C. & de Mazières, J. Grooming at a preferred site reduces heart rate in horses. *Anim. Behav.* **46**, 1191–1194. https://doi.org/10.1006/anbe.1993.1309 (1993).
71. VanDierendonck, M. C. & Spruijt, B. M. Coping in groups of domestic horses—Review from a social and neurobiological perspective. *Appl. Anim. Behav. Sci.* **138**, 194–202. https://doi.org/10.1016/j.applanim.2012.02.007 (2012).
72. Fraser, A. *The behaviour and welfare of the horse* (CABI, 2010).
73. Janczarek, I. et al. Emotional excitability and behaviour of horses in response to stroking various regions of the body. *Anim. Sci. J.* **89**, 1599–1608. https://doi.org/10.1111/asj.13104 (2018).
74. Normando, S. et al. Effect of manual imitation of grooming on riding horses' heart rate in different environmental situations. *Vet. Res. Commun.* **27**, 615–617 (2003).
75. Rochais, C. et al. Horses' tactile reactivity differs according to the type of work: The example of equine-assisted intervention. *Vet. Sci.* **10**, 130. https://doi.org/10.3390/vetsci10020130 (2023).
76. Loftus, L., Murchie, C., Asher, L. & Leach, M. C. The development of the Equine Personality Eight-factor Model (EPEM) (In Prep).
77. Tomberg, C., Petagna, M. & de Selliers de Moranville, L.-A. Horses (*Equus caballus*) facial micro-expressions: Insight into discreet social information. *Sci. Rep.* **13**, 8625. https://doi.org/10.1038/s41598-023-35807-z (2023).
78. Bornmann, T., Randle, H. & Williams, J. Investigating equestrians' perceptions of horse happiness: An exploratory study. *J. Equine Vet.* **104**, 103697 (2021).
79. Ricci-Bonot, C. & Mills, D. S. Recognising the facial expression of frustration in the horse during feeding period. *Appl. Anim. Behav. Sci.* **265**, 105966. https://doi.org/10.1016/j.applanim.2023.105966 (2023).
80. Hintze, S. & Schanz, L. Using the judgment bias task to identify behavioral indicators of affective state: Do eye wrinkles in horses reflect mood?. *Front. Vet. Sci.* **8**, 676888. https://doi.org/10.3389/fvets.2021.676888 (2021).
81. Des Roches, A. D. B., Richard-Yris, M. A., Henry, S., Ezzaouïa, M. & Hausberger, M. Laterality and emotions: Visual laterality in the domestic horse (*Equus caballus*) differs with objects' emotional value. *Physiol. Behav.* **94**, 487–490. https://doi.org/10.1016/j.physbeh.2008.03.002 (2008).
82. Leliveld, L. M., Langbein, J. & Puppe, B. The emergence of emotional lateralization: evidence in non-human vertebrates and implications for farm animals. *Appl. Anim. Behav. Sci.* **145**, 1–14 (2013).
83. Evans, L., Cameron-Whytock, H. & Ijichi, C. Eye understand: Physiological measures as novel predictors of adaptive learning in horses. *Appl. Anim. Behav. Sci.* **271**, 106152. https://doi.org/10.1016/j.applanim.2023.106152 (2024).
84. Loftus, L., Marks, K., Jones-McVey, R., Gonzales, J. L. & Fowler, V. L. Monty Roberts' public demonstrations: Preliminary report on the heart rate and heart rate variability of horses undergoing training during live audience events. *Animals* **6**, 55 (2016).
85. Scopa, C. et al. Inside the interaction: Contact with familiar humans modulates heart rate variability in horses. *Front. Vet. Sci.* **7**, 582759. https://doi.org/10.3389/fvets.2020.582759 (2020).
86. Aurich, J. et al. Effects of season, age, sex, and housing on salivary cortisol concentrations in horses. *Domest. Anim. Endocrinol.* **52**, 11–16 (2015).
87. Christensen, J. W. et al. Rider effects on horses' conflict behaviour, rein tension, physiological measures and rideability scores. *Appl. Anim. Behav. Sci.* **234**, 105184. https://doi.org/10.1016/j.applanim.2020.105184 (2021).
88. De Mira, M. C. et al. Salivary cortisol and eye temperature changes during endurance competitions. *BMC Vet. Res.* **17**, 1–12. https://doi.org/10.1186/s12917-021-02985-9 (2021).
89. Hall, C., Kay, R. & Yarnell, K. Assessing ridden horse behavior: Professional judgment and physiological measures. *J. Vet. Behav.* **9**, 22–29. https://doi.org/10.1016/j.jveb.2013.09.005 (2014).
90. Jastrzębska, E. et al. Conflict behavior in show jumping horses: A field study. *J. Equine Vet.* **57**, 116–121. https://doi.org/10.1016/j.jevs.2017.07.009 (2017).
91. Jolivald, A. *An Investigation into Physiological Correlates of Equine Personality*. Nottingham Trent University (2021).
92. Reefmann, N., Kaszas, F. B., Wechsler, B. & Gygax, L. Physiological expression of emotional reactions in sheep. *Physiol. Behav.* **98**, 235–241. https://doi.org/10.1016/j.physbeh.2009.05.017 (2009).
93. Boiten, F. A. The effects of emotional behaviour on components of the respiratory cycle. *Biol. Psychol.* **49**, 29–51. https://doi.org/10.1016/s0301-0511(98)00025-8 (1998).
94. Boiten, F. A., Frijda, N. H. & Wientjes, C. J. E. Emotions and respiratory patterns: Review and critical analysis. *Int. J. Psychophysiol.* **17**, 103–128. https://doi.org/10.1016/0167-8760(94)90027-2 (1994).
95. AWIN. AWIN Welfare assessment protocol for horses (2015).
96. Minero, M. et al. Using qualitative behaviour assessment (QBA) to explore the emotional state of horses and its association with human-animal relationship. *Appl. Anim. Behav. Sci.* **204**, 53–59. https://doi.org/10.1016/j.applanim.2018.04.008 (2018).
97. Ruet, A. et al. Horse welfare: A joint assessment of four categories of behavioural indicators using the AWIN protocol, scan sampling and surveys. *Anim. Welf.* **31**, 455–466. https://doi.org/10.7120/09627286.31.3.008 (2022).
98. Wright, J. S. & Panksepp, J. An evolutionary framework to understand foraging, wanting, and desire: The neuropsychology of the SEEKING system. *Neuropsychoanalysis* **14**, 5–39. https://doi.org/10.1080/15294145.2012.10773683 (2012).
99. Sifneos, P. E. Affect, emotional conflict, and deficit: An overview. *Psychother. Psychosom.* **56**, 116–122 (1991).
100. Maroco, J. et al. Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Res. Notes* **4**, 299. https://doi.org/10.1186/1756-0500-4-299 (2011).
101. Russell, J. A. Core affect and the psychological construction of emotion. *Psychol. Rev.* **110**, 145. https://doi.org/10.1037/0033-295X.110.1.145 (2003).
102. Burgdorf, J. & Panksepp, J. The neurobiology of positive emotions. *Neurosci. Biobehav. Rev.* **30**, 173–187. https://doi.org/10.1016/j.neubiorev.2005.06.001 (2006).
103. Kaźmierczak, M. & Nicola, S. M. The arousal-motor hypothesis of dopamine function: Evidence that dopamine facilitates reward seeking in part by maintaining arousal. *Neuroscience* **499**, 64–103. https://doi.org/10.1016/j.neuroscience.2022.07.008 (2022).
104. Kim, J., Jung, H. & Yoon, M. Relationship between plasma dopamine concentration and temperament in horses. *Domest. Anim. Endocrinol.* **83**, 106788. https://doi.org/10.1016/j.domaniend.2023.106788 (2023).
105. Kienapfel, K., Link, Y., König, V. & Borstel, U. Prevalence of different head-neck positions in horses shown at dressage competitions and their relation to conflict behaviour and performance marks. *PLoS ONE* **9**, e103140. https://doi.org/10.1371/journal.pone.0103140 (2014).
106. Valenchon, M., Lévy, F., Fortin, M., Leterrier, C. & Lansade, L. Stress and temperament affect working memory performance for disappearing food in horses, *Equus caballus*. *Anim. Behav.* **86**, 1233–1240 (2013).
107. Valenchon, M., Lévy, F., Moussu, C. & Lansade, L. Stress affects instrumental learning based on positive or negative reinforcement in interaction with personality in domestic horses. *PLoS ONE* **12**, e0170783 (2017).
108. Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150202. https://doi.org/10.1098/rsta.2015.0202 (2016).

109. Lansade, L., Neveux, C. & Levy, F. A few days of social separation affects yearling horses' response to emotional reactivity tests and enhances learning performance. *Behav. Proc.* **91**, 94–102. https://doi.org/10.1016/j.beproc.2012.06.003 (2012).

110. Fureix, C. et al. A preliminary study of the effects of handling type on horses' emotional reactivity and the human–horse relationship. *Behav. Proc.* **82**, 202–210. https://doi.org/10.1016/j.beproc.2009.06.012 (2009).

111. Löckener, S., Reese, S., Erhard, M. & Wöhr, A. C. Pasturing in herds after housing in horseboxes induces a positive cognitive bias in horses. *J. Vet. Behav.* **11**, 50–55. https://doi.org/10.1016/j.jveb.2015.11.005 (2016).

112. Vazire, S., Gosling, S. D., Dickey, A. S. & Schapiro, S. J. In *Handbook of Research Methods in Personality Psychology* (eds Robins, R. W. et al.) 190–206 (The Guilford Press, 2007).

113. Woodrum Setser, M. M., Neave, H. W. & Costa, J. H. C. The history, implementation, and application of personality tests in livestock animals and their links to performance. *Appl. Anim. Behav. Sci.* **268**, 106081. https://doi.org/10.1016/j.applanim.2023.106081 (2023).

114. Weiss, A. & Robinson, L. Subjective well-being, happiness, and personality in animals 111–122 (CABI, 2020). https://doi.org/10.1079/9781786393401.0111.

115. Williams, E., Chadwick, C. L., Yon, L. & Asher, L. A review of current indicators of welfare in captive elephants (*Loxodonta africana* and *Elephas maximus*). *Anim. Welf.* **27**, 235–249. https://doi.org/10.7120/09627286.27.3.235 (2018).

116. Kuppens, P. et al. The relation between valence and arousal in subjective experience varies with personality and culture. *J. Pers.* **85**, 530–542 (2017).

117. Stratton, R. Assessment of positive emotion in horses. PhD thesis, Massey University, University of New Zealand (2022).

118. Leiner, L. & Fendt, M. Behavioural fear and heart rate responses of horses after exposure to novel objects: Effects of habituation. *Appl. Anim. Behav. Sci.* **131**, 104–109. https://doi.org/10.1016/j.applanim.2011.02.004 (2011).

119. Stull, C. L. & Rodiek, A. V. Effects of cross-tying horses during 24 h of road transport. *Equine Vet. J.* **34**, 550–555. https://doi.org/10.2746/042516402776180214 (2002).

120. Birke, L. et al. Horses' responses to variation in human approach. *Appl. Anim. Behav. Sci.* **134**, 56–63. https://doi.org/10.1016/j.applanim.2011.06.002 (2011).

121. Górecka, A., Bakuniak, M., Chruszczewski, M. H. & Jezierski, T. A. A note on the habituation to novelty in horses: handler effect. *Anim. Sci. Pap. Rep.* **25**, 143–152 (2007).

122. Zhao, D., Wang, Y. & Li, B. Eye preference for emotional stimuli in Sichuan snub-nosed monkeys. *Folia Primatol.* **91**, 630–642. https://doi.org/10.1159/000509442 (2020).

123. McDonnell, S. M. & Poulin, A. Equid play ethogram. *Appl. Anim. Behav. Sci.* **78**, 263–290. https://doi.org/10.1016/S0168-1591(02)00112-0 (2002).

124. Sigurjónsdóttir, H., Van Dierendonck, M. C., Snorrason, S. & Thórhallsdóttir, A. G. Social relationships in a group of horses without a mature stallion. *Behaviour* **140**, 783–804. https://doi.org/10.1163/156853903322370670 (2003).

125. Ninomiya, S., Kusunose, R., Sato, S., Terada, M. & Sugawara, K. Effects of feeding methods on eating frustration in stabled horses. *Anim. Sci. J.* **75**, 465–469. https://doi.org/10.1111/j.1740-0929.2004.00214.x (2004).

126. RedGorilla. *Gorilla Tub Mini Shallow 5L.* https://www.redgorilla.red/products/gorilla-tub-mini-5l?variant=47370075570486 (2024).

127. McDonnell, S. M. *The Equid Ethogram: A Practical Field Guide to Horse Behavior* (Eclipse Press, 2003).

128. Waring, G. H. *Horse Behaviour. The behavioral Traits and Adaptations of Domestic and Wild Horses, Including Ponies* 2nd edn. (William Andrew, 1983).

129. Vasconcellos, A. S. et al. Training reduces stress in human-socialised wolves to the same degree as in dogs. *PLoS ONE* **11**, e0162389 (2016).

130. Puppe, B., Ernst, K., Schön, P. C. & Manteuffel, G. Cognitive enrichment affects behavioural reactivity in domestic pigs. *Appl. Anim. Behav. Sci.* **105**, 75–86 (2007).

131. Martin, P. & Bateson, P. *Measuring Behaviour: An Introductory Guide* (Cambridge University Press, 2007).

132. Malik, M. et al. Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation* **93**, 1043–1065. https://doi.org/10.1161/01.CIR.93.5.1043 (1996).

133. Kędzierski, W., Strzelec, K., Cywińska, A. & Kowalik, S. Salivary cortisol concentration in exercised thoroughbred horses. *J. Equine Vet. Sci.* **33**, 1106–1109. https://doi.org/10.1016/j.jevs.2013.04.011 (2013).

134. Alhajj, M., Zubair, M. & Farhana, A. *Enzyme Linked Immunosorbent Assay* (StatPearls, 2023).

135. Kim, S.-M. & Cho, G.-J. Validation of eye temperature assessed using infrared thermography as an indicator of welfare in horses. *Appl. Sci.* **11**, 7186. https://doi.org/10.3390/app11167186 (2021).

136. Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O. & Karjalainen, P. A. Kubios HRV—Heart rate variability analysis software. *Comput. Methods Programs Biomed.* **113**, 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024 (2014).

137. Hänninen, L. & Pastell, M. CowLog: Open source software for coding behaviors from digital video. *Behav. Res. Methods* **41**, 472–476. https://doi.org/10.3758/BRM.41.2.472 (2009).

138. Pastell, M. CowLog—Cross-platform application for coding behaviours from video. *J. Open Res. Softw.* **4**, 15. https://doi.org/10.5334/jors.113 (2016).

139. Masko, M., Domino, M., Lewczuk, D., Jasinski, T. & Gajewski, Z. Horse behavior, physiology and emotions during habituation to a treadmill. *Animals* **10**, 921. https://doi.org/10.3390/ani10060921 (2020).

140. Bennett-Wimbush, K., Amstutz, M. & Duthie, M. Effects of overcheck use on stress parameters and welfare implications in driving horses. *J. Appl. Anim. Welf. Sci.* **23**, 83–94. https://doi.org/10.1080/10888705.2019.1594229 (2020).

141. Contalbrigo, L. et al. Equine-assisted interventions (EAIs) for children with autism spectrum disorders (ASD): Behavioural and physiological indices of stress in domestic horses (*Equus caballus*) during riding sessions. *Animals* **11**, 1562. https://doi.org/10.3390/ani11061562 (2021).

142. Cerasoli, F. et al. Assessment of welfare in groups of horses with different management, environments and activities by measuring cortisol in horsehair, using liquid chromatography coupled to hybrid orbitrap high-resolution mass spectrometry. *Animals* **12**, 1739. https://doi.org/10.3390/ani12141739 (2022).

143. Kremer, L., Klein Holkenborg, S. E. J., Reimert, I., Bolhuis, J. E. & Webb, L. E. The nuts and bolts of animal emotion. *Neurosci. Biobehav. Rev.* **113**, 273–286. https://doi.org/10.1016/j.neubiorev.2020.01.028 (2020).

144. Marinier, S. & Alexander, A. The use of a maze in testing learning and memory in horses. *Appl. Anim. Behav. Sci.* **39**, 177–182 (1994).

145. Dalla Costa, E., Murray, L., Dai, F., Canali, E. & Minero, M. Equine on-farm welfare assessment: A review of animal-based indicators. *Anim. Welf.* **23**, 323–341. https://doi.org/10.7120/09627286.23.3.323 (2014).

146. McGreevy, P. D., Oddie, C., Burton, F. L. & McLean, A. N. The horse-human dyad: Can we align horse training and handling activities with the equid social ethogram?. *Vet. J.* **181**, 12–18. https://doi.org/10.1016/j.tvjl.2009.03.005 (2009).

147. Fortin, M. et al. Emotional state and personality influence cognitive flexibility in horses (*Equus caballus*). *J. Comp. Psychol.* **132**, 130–140. https://doi.org/10.1037/com0000091 (2018).

148. Lansade, L. et al. Personality and predisposition to form habit behaviours during instrumental conditioning in horses (*Equus caballus*). *PLoS ONE* **12**, e0171010. https://doi.org/10.1371/journal.pone.0171010 (2017).

149. Fureix, C., Jego, P., Henry, S., Lansade, L. & Hausberger, M. Towards an ethological animal model of depression? A study on horses. *PLoS ONE* **7**, e39280. https://doi.org/10.1371/journal.pone.0039280 (2012).

150. R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2023).
151. Voudouris, K. et al. Direct human-AI comparison in the animal-AI environment. *Front. Psychol.* **13**, 711821. https://doi.org/10.3389/fpsyg.2022.711821 (2022).
152. Jöreskog, K. G. *Handbook of Multivariate Experimental Psychology* 207–230 (Springer, 1988).
153. Cleophas, T. J. Carryover bias in clinical investigations. *J. Clin. Pharmacol.* **33**, 799–804. https://doi.org/10.1002/j.1552-4604.1993.tb01954.x (1993).

Acknowledgements

Author contributions

L.L. contributed to the conception and design of the study, collected the data, analysed and interpreted the data and drafted the article. A.N. contributed to the conception and design of the study and to data collection within the pilot study. M.L. contributed to the conception and design of the study and revision of the article. L.A. contributed to the conception and design of the study, analysis of data and revision of the article.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98034-8.

Correspondence and requests for materials should be addressed to L.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.