Is the judgment bias test a good tool to assess the quality of horse management?

Published Version:
Marliani, G., Balboni, A., Tiberi, C., Malavasi, R., Gardini, A., Accorsi, P.A. (2022). Is the judgment bias test a good tool to assess the quality of horse management?. JOURNAL OF VETERINARY BEHAVIOR, 58(December), 62-69 [10.1016/j.jveb.2022.11.002].

Availability:
This version is available at: https://hdl.handle.net/11585/911658 since: 2023-05-28

Published:
DOI: http://doi.org/10.1016/j.jveb.2022.11.002

Is the judgment bias test a good tool to assess the quality of horse management?

Giovanna Marliani*[a], Alessia Balboni[a], Carlotta Tiberi[a], Rachele Malavasi[b], Aldo Gardini[c], Pier Attilio Accorsi[a]

[a]Department of Veterinary Medical Sciences, University of Bologna, Ozzano dell'Emilia (BO), Italy

[b]School of Ethical Equitation, Moncigoli di Fivizzano, +3498551997, Italy

[c]Department of Statistical Sciences P. Fortunati, University of Bologna, Bologna, Italy

Abstract

Animal welfare should include the possibility of animals experiencing positive emotions. Emotions influence the cognitive process, and judgment bias tests (JBTs) are employed in different species, to assess the optimistic or pessimistic expectation of an individual and its affective state. Only recently the JBTs have been applied to horses. This research aims to investigate the relationship between a spatial JBT and hypothalamic-pituitary-adrenal axis (HPA) chronic and acute activation in forty-one animals hosted in different kinds of environments: traditional stables (TS), natural boarding (NB), and ethological stable (ES). Fecal (FC) and horsehair (HC) cortisol concentrations were quantified for each subject through Radio-Immuno-Assay (RIA). Body condition score (BCS), as an indirect index of animal motivation towards food, and personality traits were measured to explore their possible influence on JBT results. Horses had to distinguish a positive position (P), where a bucket full of food was positioned, from a negative one (N), with an empty bucket. Then, 3 intermediate positions (Near Negative-NN, Near positive-NP, and Medium-M) with an empty bucket were presented to the subject one at a time. Only 20 subjects out of 41 completed the JBT and were included in the statistical analysis, and both BCS and P position, whether at the right or the left of the subject, seems to have influenced the inclusion rate. Only the ES group registered a significantly lower score in NN, suggesting a more optimistic affective state, whereas NB and TS did not significantly differ in their responses. Despite this, horses from NB recorded higher FC concentration than TS subjects during all the phases of the test, but lower HC levels, which could suggest a generally lower level of chronic stress but its interpretation presents several confounding factors. These results put into question whether

JBT is indeed a good test to monitor the quality of the management, as it does not seem to reflect the chronic physiological state of the animals and could be influenced by a state of acute stress, caused by the test procedure. Due to these confounding factors, this procedure should be accompanied by other indicators. Finally, to include more animals and exclude possible biases, the structure of the JBT and the employment of food as a reward should be evaluated considering the peculiarities of the species and individual motivations.

Keywords Cognitive test; Equine; Emotions; Welfare; Cortisol; Stress

Introduction

The consideration of emotional aspects in animal welfare science has increasingly gained importance (Broom, 2003). According to the 5 Domains Model, animal management should reduce situations that negatively affect animals and promote the experience of positive emotions (Mellor and Beausoleil, 2015; Mellor et al., 2020). However, the investigation of welfare is often focused on external circumstances, while affective states are difficult to explore (Duncan, 2006; Mellor et al., 2020).

Cortisol is commonly employed in animal welfare studies as a physiological indicator of stress. Nevertheless, the increase of this glucocorticoid can occur as a normal non-stress-physiological response or can be caused by both negative and positive situations, such as pain, anxiety, and fear as well as excitement or pleasure (Koolhaas et al., 2011; Ralph and Tillbrook, 2016). Therefore, it is always important to contextualize and try to understand what potentially has caused the activation of the hypothalamic-pituitary-adrenal axis (HPA), which leads to the increase of cortisol (Ralph and Tillbrook, 2016). Feces and horsehair are used as non-invasive methods for the determination of cortisol concentrations (Palme, 2012; Duran et al., 2017). While feces evidence the increase of plasmatic cortisol according to the animals' metabolism (in horses after about 24 hours) (Mostl et al., 1999), horsehair reflects the individual history, reflecting chronic stress conditions (Duran et al., 2017).

However, stress and, consequently, cortisol levels can be influenced by both emotional and physical stressors. Therefore, to investigate emotions in non-vocal animals, research relies on indirect indications of how the animal perceives the environment, considering that emotional states can influence memory, attention, and judgment (Duncan, 2006; Mendl et al., 2009). In particular, Judgment bias tests (JBTs), modulated by human sciences (Bethell, 2015), have been employed in animal research to investigate how the environment or routine procedures affect animals' affective states (Burman et al., 2009, Neave et al., 2013; Daros et al., 2014). In animal research, judgment bias represents the animal's response to a neutral stimulus and, evaluating if the individual expects from this stimulus either a positive or a negative outcome, it defines the subject as optimistic or pessimistic (Mendl et al., 2009).

During recent times, the JBT has been applied in horses (Freymond et al., 2014; Löckener et al., 2016). Domestic horses' welfare and mental health depend on both their relationship with human beings and their management (Hausberger et al., 2009; Freymond et al., 2014). Methods of horses' management are varied and differ in terms of freedom of movement, the possibility of interaction with conspecifics, and the kind of feeding and its routine. In some countries, domestic horses are traditionally housed in individual stalls, where they spend most of their time (Yarnell et al., 2015). This kind of management is considered practical for humans because, for example, it is easier to detect any diseases, control the consumption of water and food, and manage the horse for various routine operations (McGreevy, 2012; Ruet et al., 2019). The natural boarding system of the horse tries to reproduce the living conditions of feral equines. Horses usually live in a group in paddocks designed to stimulate movement, exploration, relationship with conspecifics, and body care (McIlwraith and Rollin, 2011; Saba et al., 2013). Finally, in Italy, there is a new kind of facility, called Ethological Stable, that has intermediate characteristics of the traditional stable and natural boarding system, and it's characterized by bigger stalls than traditional ones and separated by grates with windows, which allow the horses to see and touch each other (Marliani et al., 2021). Whichever management method is considered, it has a profound effect on the welfare of the individual (Yarnell et al., 2015; McGreevy, 2012; Ruet et al., 2019), and JBT could be helpful for the assessment of the welfare impacts of different housing systems, but its application in horses requires further investigation.

This research evaluated horses under 3 different management systems: traditional stable, natural boarding, and ethological stable. Each subject underwent a JBT, fecal and horsehair sampling, and an evaluation of personality traits and Body Condition Score (BCS). We tested if the type of management can influence horses' judgment bias. We also investigated the influence of other independent variables (personality, age, and BCS) on judgment bias, and we examined if JBT results correlated to cortisol levels found in feces and horsehair.

Material and Methods

Stables

Natural Boarding System (NB)

This system hosted 21 horses, living 24/7 together in a large paddock of about 6 ha. The paddock was composed of heterogeneous spaces with large areas and corridors, huts and muddy and sandy soil areas (according to season). Trunks and objects, where animals can rub, were scattered within the paddock, to stimulate the movement of animals, and the entire area was arranged in the shape of a labyrinth. Hay and water were available ad libitum and horses were occasionally allowed to access grassy lawns. This system was frequented daily by at least 1 person, from a few hours to the entire day. Most of the horses were used for occasional horse trekking and horse-riding activities in the sandy field (maximum 2 hr/day) by riders of varying ages and experience. Only 2 of the horses included in the statistical analysis were not involved in these activities. Horses were barefoot and ridden without a bit.

Traditional Stable (TS)

This facility housed 55 horses during the study. The stable was organized in 2 parallel corridors that had separate rows of individual stalls (3.00 m x 3.50 m). Each stall had wood chip bedding, a sliding door and a window that allowed the horse to look out into the corridor, facilitating visual contact with the other horses in the stable. The facility had individual paddocks (10.00 m x 10.00 m; or 20.00 m x 20.00 m) where horses were pastured for a different lag of time that varied depending on weather conditions and the owner's decisions, not every day and only during the daytime. Hay was provided at 7.00 am and 4.00 pm and a supply of mixed grains were provided at 8.00 am and 8.00 pm. Most of the horses housed in this stable were involved in sporting activities, in particular jumping, dressage, and vaulting with beginners and competitors. The stable was frequented every day and had a carousel, a round pen, and some sand courts.

Ethological Stable (ES)

The system was the same as described by Marliani et al., 2021. During the study, the stable hosted 11 horses. The stable included 12 stalls, called Big Box© (8.00 m x 5.00 m), arranged in rows of 6. Each stall was divided into 2 sections of 20m2: the front part consisted of high-density impact rubber mats that provided the horse with rigid support; the rear part consisted of a soft and semi-permanent element of river sand. This system of double litter allowed the horse freedom of movement and choice of the ground on which to rest. The walls of the Big Box© were of metal gratings, with windows between stalls, so that subjects were able to see and touch each other. There was a paddock of 1.50ha organized with large areas and corridors, and 6 paddocks of 2.00ha each. Horses, individually or in a small group, stayed in paddocks all day and were brought into the stall for the evening, or in case of bad weather. Hay was available throughout the day and cereals (barley, oats, and corn), apples, and carrots were provided at 7.00 a.m. and 7.00 p.m., according to a personalized food plan. There were also a round sand pen and a covered sand rectangular field. Horses were engaged principally in flatwork, country riding, and walks, and they were not used for competitions.

Subjects

Forty-one adult horses were involved in the study: 16 subjects from NB, 14 horses from TS, and 11 subjects from ES (Table 1). All horses had been resident in each stable for at least the last 3 months. The history of each subject and the characteristics of the stables were collected using 2 questionnaires.

Body Condition Score (BCS) and Personality Evaluation

An evaluation of BCS was made for every horse. This is a nutrition assessment system, in which a score is assigned based on the visual (rib and bone prominence) and manual evaluation of the panniculus adipose in 6 main regions: neck, shoulder, withers, thorax, lumbar region, and tail attachment. First of all, the amount of fat present is visually evaluated, assessing the visibility or not of bone and rib prominences. The score is from 1 to 9, where 1 corresponds to excessive emaciation and 9 is a condition of excessive nutrition (Henneke et al., 1983). The BCS was collected and considered in the statistical analysis for all the horses, except for 3 (2 NB horses who suffered from Cushing and 1 ES individual).

To assess the personality of the subjects involved in the study, a translated version of the HPQ (Horse Personality Questionnaire) (Lloyd et al., 2007) was used. The HPQ was completed by the horse's owner or caregiver. For each of the 25 items, they were asked to assign a score on a Likert scale from 1 to 7, where 1 indicates that the trait was not at all manifested, while 7 corresponds to the extreme expression of the character trait. The data were analyzed according to Lloyd et al. (2007), obtaining a score for each of the 6 personality components (Dominance, Anxiousness, Excitability, Protection, Sociability, and Inquisitiveness).

Judgment Bias Test

The protocol employed in the study was a modification of those proposed by Freymond et al. (2014) and Henry et al. (2017). It was a Go/No-Go JBT with spatial stimuli. Black buckets were employed for the test, to prevent the horses seeing the contents. In each stable, the test was conducted in a fenced arena with a sandy substrate, familiar to the subjects and sufficiently far from disturbances.

In the arena, a start-line and a semicircular stop-line were drawn on the ground using the wood-chip, to allow the experimenter to see better when the horses passed the line to stop the stopwatch. The test was carried out by 2 trained experimenters. Experimenter 1 (E1) was in charge of preparing the bucket and placing it behind the finish line, 9 meters from the start line (Figure 1). Experimenter 2 (E2) had the task of leading the horse into the test arena and placing it behind the start line. Only when E1 had placed the bucket, the E2 release the horse and take a step to the side. Once the horse finished the trial, E2 retrieved the horse and brought it back to the start line. The trial was completed when the horse passed the finish line with one of the forelegs, or after 60 seconds, which was the maximum latency considered. E2 measured the time taken by the horse to reach and pass the stop line. While the horse completed the trial, the experimenters remained stationary and directed their gaze towards the ground to avoid influencing the behavior of the horse. The experimenter who led the horse was always the same for the same horse to get the horse used to the same person and minimize biases due to the possible stress caused by different experimenters. The JBT was divided into 4 consecutive days: habituation, training, and 2 sessions of the test.

Habituation

The first day was employed for the phase of habituation, during which horses got used to the setting. During this phase, 1 horse at a time was led into the arena by an experimenter, then the second experimenter positioned the bucket with a reward, which consisted of a small amount of oats, at the positive position (P). Oats was chosen as a reward because usually horses are highly motivated by them. The habituation was completed when horses, released at the start line, independently approached the bucket 3 consecutive times. The horses were randomly allocated to either the left or right position of the test apparatus.

Training

For the training phase the negative position (N), corresponding to an empty bucket, was set at the opposite in respect of the positive (P), 7.20 meters apart. The 2 positions were presented 1 at a time, and the same position was presented no more than twice in a row. The training always finished with P. E2 recorded the latency of each subject to reach and pass the finish line. If the horse didn't move towards the bucket in 60 seconds, E1 went towards the horse, took it, and led it again to the start position and the trial was considered completed. The training ended when the latencies to reach the last 3 positive buckets were lower than that to reach the last 3 negative ones (Mendl et al., 2010). Following the protocol suggested by Henry et al. (2017), each training session was composed of 3 positive positions and 2 negative locations, and finished always with the positive one. The minimum number of sessions employed was two. If the horse failed to distinguish the 2 positions after 5 training sessions, it was excluded from the JBT.

Testing

The day after the training, the testing phase was performed in 2 consecutive days. During the test, 3 intermediate positions, 1 at a time, between N and P were presented: NN (Near-Negative), M (Middle), and NP (Near-Positive). These were placed 1.80 meter from each other and the bucket was always empty. Before the beginning of the test, N and P were presented as a reminder of the training, a maximum of 6 times. Each session followed the scheme proposed by Henry et al. (2017), where ambiguous locations were preceded alternately by positive and negative locations:

• first session (first day testing phase) P-N-NP-P-N-M-P-N-NN-P

• second session (second day of testing phase) N-P-M-N-P-NN-N-P-NP

The second session always ended with an N position where the bucket was full of food, to exclude horses employed their sense of smell during the test (Mendl et al., 2010).

The test sessions were videotaped using a Sony HDR-CX240E camera placed on a tripod outside the arena. All videos were analyzed using the software Solomon Coder, to record precisely the time that horses employed to reach buckets.

To avoid biases caused by differences in baseline running speeds, due to size and/or age of individuals, raw latencies recorded to reach the intermediate positions were transformed into scores, according to the formula proposed by Mendl et al. (2010).

Adjusted latency = (mean latency to ambiguous location −mean latency to P) / (mean latency to N −mean latency to P)

This formula returns 0 for the P and 100 for N.

Collection of Fecal and Horsehair Samples

Fecal samples were collected immediately after defecation, for a total of 4 samples for each horse that completed the test. The collection was made during:

1. the habituation (P1), which was representative of the baseline cortisol concentration before the beginning of the whole procedure;

2. the training (P2), which was representative of cortisol level during habituation;

3. the first session of the test (P3), which was representative of training;

4. the second session of the test (P4), representative of the first day of the test.

A total of 86 fecal samples were collected. Some samples were missing for the NB group. One horse was excluded because he suffered from Cushing disease, which alters cortisol concentration, and some could not be collected for other reasons (1 sample during the habituation; 4 samples during the training, and 1 sample for the first session of the test). Each fecal sample was placed in a non-sterile container identified with date, time, subject name, refrigerated to +4°C, transported to the laboratory and frozen to – 20°C.

Horsehair samples were collected from the base of the mane of each horse who completed the JBT and stored at room temperature in identified non-sterile containers, for a total of 20 samples. The sample of the horse who suffered from Cushing was excluded from the statistical analysis.

Cortisol Assay

Cortisol concentrations were determined by Radio Immuno-assays (RIAs). Cortisol was extracted from fecal specimens (500 mg, wet weight) with methanol-water solution (v/v 4:1) and ethyl ether, whereas for horsehairs 60 mg of trimmed horsehairs (1-3 mm) of a subject were put in a glass vial with 5 ml methanol (Accorsi et al., 2008). All samples were dried under an air-stream suction hood at 37°C and the dry-residue was dissolved into phosphate-buffered saline (PBS) 0.05 M, pH 7.5. Cortisol metabolites assay in both feces and horsehair were carried out according to Tamanini et al. (1983). The cortisol RIA was performed using an antiserum to cortisol-21-hemisuccinate-BSA (anti-rabbit), at a working dilution of 1:20 000 and 3H-cortisol (30 pg/tube vial) as a tracer. Validation parameters of analysis were: sensitivity 0.19 pg/mg, intra-assay variability 5.9%, inter-assay variability 8.7%. Radioactivity was determined using a liquid scintillation β counter and a linear standard curve, ad hoc designed by a software program (Motta and Degli Esposti, 1981).

Statistical Analysis

To analyze the collected data, non-parametric statistical methods were employed, given that the small size of the sample did not guarantee the reliability of the normality assumption. In particular, to assess the presence of differences among the management, the Kruskal-Wallis test with Dunn's adjustment for multiple comparisons was adopted. To test for the presence of correlation between numerical variables, Spearman's correlation coefficient was considered. To assess the influence of the position of nominal and numerical variables on the inclusion or the exclusion of the horses from the test, we employed a Chi-squared test and a Mann-Whitney test. To compare the fecal cortisol concentration during the different phases of the test we employed a Friedman test, with a Wilcoxon pairwise as post-hoc test. Statistical significance was set at $P < 0.05$. The whole statistical analysis was carried out in the R environment (R Core Team, 2021).

Results

Task Acquisition

Of the 41 horses considered, only 20 horses completed the JBT and were considered for the statistical analysis (NB $n = 10$; ES $n = 4$; TS $n = 6$). Eleven horses did not pass the habituation phase (NB $n = 6$; ES $n = 1$; TS $n = 4$) and 4 subjects the training (ES $n = 4$). Moreover, 6 did not maintain a significantly different latency between P and N during the reminders before the test phase (ES $n = 2$, TS $n = 4$). The rate of excluded animals was similar for the 3 types of management.

All the horses included employed the same time to reach N when there was both the empty and the full bucket (last position of the second test phase).

Twenty-two horses had the positive bucket on their left side, 19 on their right. Considering this variable and the entire group in a chi-squared test, there was a significantly higher probability that the horses who had the P location on their left side was excluded from the test ($\chi2 = 5.47$, df = 1, P-value = 0.02). Indeed, considering the subjects included in the test, 13 had P on their right side and 7 on their left.

The BCS of the excluded horses was significantly (W = 104; r = -0.359; 95%C.I. [-0.614, -0.109]; P-value = 0.02) lower (Median ±IQR: 4 ± 1.5, n = 19) than that of the included subjects (5 ± 2.0, n = 19) (Figure 2).

JBT Scores

Comparing the 3 different stables, scores recorded in NP ($\chi2$= 4.2957, df = 2, P-value = 0.12) and M ($\chi2 = 1$, df = 2, P-value = 0.61) did not significantly differ between the groups, but there was a significant difference for NN position ($\chi2$= 8.3082, df = 2, P-value = 0.02). Indeed, the horses of ES had an NN score significantly lower (6.46 ± 5.94) than subjects of NB (100.00 ± 46.50; P-value = 0.01) and than those of TS (72.34 ± 35.75; P-value = 0.04) (Figure 3).

Considering the entire group of horses, the scores recorded in NP (Median ± IQR: 10.24 ± 44.63), M (2.88 ± 17.34), and NN locations (71.09 ± 67.00) were significantly different, according to the Kruskal-Wallis test with Dunn's post-hoc test ($\chi2$= 18.7236, df = 2, P-value < 0.001). In particular, NN was significantly different from NP (P-value < 0.01) and M (P-value < 0.001), but NP and M did not differ (P-value = 0.34). No correlation was found between the scores recorded for the intermediate positions and personality traits and age, but there was a significant negative correlation between the score in NP and BCS (S = 1907.1, P < 0.01, rho = -0.67). Body condition score differs between groups ($\chi2$= 9.0149, df = 2, P-value = 0.01) and it was significantly higher in NB (6.22 ± 1.72) horses than in TS ones (3.83 ± 0.75, P-value = 0.01), while ES subjects recorded an intermediate value (5 ± 0), and did not differ neither from TS (P-value = 0.14) and NB (P-value = 0.64).

Cortisol Results

Fecal cortisol (FC) baseline level (P1) and its concentration during the habituation phase (P2), training (P3) and the first session of the test (P4) were determined from the fecal samples collected during the 4 days of the JBT. Comparing the FC concentration in P1, P2, P3 and P4, it was not significantly different in both TS groups ($\chi2$=2, df = 3, P-value = 0.57) and ES ($\chi2$= 8.1, df = 3, P-value = 0.051).

The FC baseline concentration was not significantly different between the 3 groups of horses ($\chi2$= 0.56, df = 2, P-value = 0.75). However, during habituation ($\chi2$= 6.67, df = 2, P-value = 0.04), training ($\chi2$= 11.59, df = 2, P-value < 0.001) and the first day of the test ($\chi2$= 11.97, df = 2, P-value < 0.001), subjects from NB registered a significantly higher concentration of FC than the TS group. The ES group showed a significantly higher concentration than TS horses in P3 (P-value = 0.03) and, with a significant tendency, in P4 (P-value = 0.07). (Table 2).

Considering horsehair cortisol (HC), its concentration resulted significantly different in the 3 groups (chi-squared = 13.93, df = 2, P-value < 0.001). It was higher in TS horses than NB individuals (1.03 ± 0.16 pg/mg; 0.21 ± 0.19 pg/mg; P < 0.001), while the ES group, which had intermediate values (0.38 ± 0.14 pg/mg), did not differ significantly from either of the others (ES vs. NB, P-value = 0.20; ES vs. TS, P-value = 0.14). In addition, as suggested by Heimbürge et al., 2019, we analyzed the influence of age and sex. We found a significant difference for sex (W = 9, r = -0.54, P-value = 0.01), and female (0.25 ± 0.12) had HC levels than geldings (0.52 ± 0.66). In addition, HC concentration was significantly positively correlated with age (S = 615.39, P-value = 0.047; rho = 0.46), but the groups did not differ significantly for age ($\chi2$=5.1301, df = 2, P-value = 0.08) . No significant results were found considering the influence of sex (W = 27, P-value = 0.96) and age (S = 1025.2, P-value = 0.82, rho = -0.05) on the fecal cortisol baseline (P1).

Discussion

The concept of animal welfare includes the possibility of animals experiencing positive emotions, whereas negative emotional and experiential states, such as fear, are sufficient elements to compromise it (Dawkins, 2006). The primary function of emotions is to guide individual decisions and they are strictly related to the cognitive process (Paul et al., 2005; Roelofs et al., 2016; Mendl et al., 2009). In particular, the judgment bias indicates how an individual judges a certain situation in dependence of their emotional state (Paul et al., 2005).

The evaluation of judgment bias is considered useful for comparing the emotional state of individuals of the same species living in different environmental situations (Bethell, 2015). This research aimed to understand the usefulness of JBT in the evaluation of welfare in horses in different housing and management systems by comparing results between JBT and chronic HPA activation. We also considered other independent variables, such as BCS, age and personality traits, to determine their influence on the outcome of the test.

Only 20 on 41 horses completed all the phases of JBT, which poses serious doubts about the possibility to generalize the results (Roelofs et al., 2016). However, as the study of JBT in horses is relatively recent and the difficulty of this species in learning and memorizing the spatial discrimination between P and N has already been reported (Hintze et al., 2018), we consider this study to be a useful addition to the current literature. The learning processes of the subjects across the trial was not influenced by the type of management, similar to what is reported by Henry et al. (2017). However, we identified 2 other factors that could have influenced the exclusion rate: the position of the positive bucket and the BCS. As for the position of the bucket, we cannot exclude a phenomenon of lateralization. Lateralization in vertebrates is the different specialization of the 2 cerebral hemispheres and it can be observed both in motor and sensorial (olfaction and sight) activities (Vallortigara and Rogers, 2005). Horses have eyes in a lateral position, with decussation of 80-90% of optic fibres (Brooks et al., 1999). Similar to other species, they usually use the right eye, so the left hemisphere, to investigate novel objects, and the left eye to evaluate objects with an emotional negative valence (De Boyer Des Roches et al., 2008). The positioning of the positive bucket on their left side could have negatively influenced the value that the subjects have assigned to the object. Moreover, the positive position was the first and only position presented to the horse during the entire habituation phase. Another factor that seems to affect the inclusion/exclusion rate is the BCS of the horses. Indeed, the excluded subjects have a BCS significantly lower than the included ones. BCS is an indirect measure of the adequate nutritional intake and health state of the individuals, and a potential index for the presence of pathologies. A healthy equid should have a BCS between 4 and 6 (Henneke et al., 1983; Yngvesson et al., 2019) and in this range fall the median values registered in all the 3 stables. However, if we consider the BCS as a potential indicator of motivation towards food (Henry et al., 2017), horses with a lower BCS could be less motivated and this can negatively influence the learning of a discrimination task based on food rewards (McCall, 1989). In addition, we should consider the different management and the different activities the horses are used for. Indeed, in the 3 management we noticed that NB horses had a significant higher BCS than TS. Not only feeding routine, health status and environment can influence BCS, but also the type of exercise (Christie et al., 2006; Zoller et al., 2019). Therefore, the differences that we found in the 3 systems can be also due to the differences of equestrian activities in which horses were involved.

The 20 horses included in the test obtained different scores for each intermediate position, demonstrating that they can differentiate NP, M, and NN. All individuals included showed the same latency to reach N also in the last trial when the bucket was filled, demonstrating that they did not use their sense of smell to deal with the JBT (Mendl et al., 2010).

The first of our predictions was that independent variables, such as body condition score (BCS), age and personality traits, do not influence the JBT results. No correlation was found between JBT scores and both age and personality traits. Regarding personality traits, this result agrees with Lalot et al. (2017) that did not evidence any correlation between personality traits and an optimistic bias in canaries. On the contrary, in another study by Barnard et al. (2018), it was found that dogs with higher sociability and excitability approached faster the ambiguous probes of a spatial JBT than those characterized by high separation-related anxiety and dog-directed fear. Therefore, the relationship between personality and JBT seems to be complex as its outcomes are influenced not only by personality traits but also by transient affective states (Whittaker and Barker, 2020), which are difficult, if not impossible, to describe in a non-vocal species. Moreover, the HPQ was validated in English and we employed a translated version of the HPQ, so we cannot exclude the presence of possible bias due to the use of a different language. Instead, a negative correlation was found between BCS and the score in NP, so the latency to reach the near-positive position was lower for horses with a higher BCS. This result is in agreement with the previous result about the inclusion rate and confirms the fact that a different motivation among individuals can influence the JBT result.

Secondly, we predicted that different management styles would have affected JBT results. Only the ES group showed a significantly lower latency to reach NN, suggesting that those horses had greater anticipation of a positive outcome at that location, and thus experienced a general more positive affective state. NN is usually associated to negative outcome and a slower latency to reach this probe location is usually interpreted as index of fear and anxiety, so of a pessimistic affective state (Barnard et al., 2018). Indeed, our results are similar to those of Karagiannis and colleagues (2015) that found that dogs suffered from separation related problems approach significantly slower NN probes but not the other locations. However, this result should be considered with caution, because only 4 of 11 horses completed the JBT in the ethological stable. Contrary to our predictions, NB and TS did not score differently to reach intermediate positions, although these 2 management systems are extremely different. It cannot be excluded that the manipulation by strangers of NB horses and, overall, the separation from their herd caused a mild state of distress, which could have altered the results of JBT. This hypothesis can be partially confirmed by the analysis of fecal cortisol, which evidenced an increasing level of glucocorticoids in the feces (representative of daily levels of stress) and a significantly higher FC concentration in NB in respect to TS subjects during the JBT procedure. According to Roelofs et al. (2016), involving social animals individually in JBT could increase stress perception. Indeed, it is well known that social isolation induces stress in most gregarious species, manifested through behavioral and physiological changes such as increasing vocalization, locomotion, heart rate, and plasma cortisol levels (Mal et al., 1991; Wolff, Hausberger and Le Scolan, 1997; Lansade and Bouissou, 2008), and the horse is a social animal that forms preferential bonds within the herd to guarantee social stability (Beaver, 2019). Therefore, the procedure itself could have been stressful for NB horses, which are also usually less used to being manipulated by strangers, and this can have determined that they did not differ from the TS group in their latency to reach intermediate positions. All these elements suggest that JBT may not be reliable to evaluate the quality of the management style for these individuals because it is affected by contingent confounding factors.

Finally, we predicted that acute and chronic activation of the HPA axis would be different according to the management system and this would be reflected in the JBT results. As for the acute activation, the fecal cortisol results are partly in accordance with the JBT results, as explained in the previous paragraph. In any case, we recommend collecting fecal samples at least 2 days after the end of the JBT procedure, to check the level of cortisol also during the second day of testing and at the return to the baseline. Fecal cortisol concentrations help us to measure not only cortisol base line, but also animal's stress response, avoiding the invasiveness of blood collection. It reflects an average of the plasma cortisol secreted and metabolized by the animal over the course of a species-specific period (in horses about 24 hours) (Mostl et al., 1999; Sheriff et al., 2011). Therefore, the daily collection of fecal cortisol allowed us to monitor the stress of animals possibly caused by the entire procedure of JBT, considering also that it could be a stressful disruption of their daily routine.

We obtained surprising results when we considered horsehair cortisol concentration, whose results, if matching with those of JBT, would have reinforced the role of JBT as a useful method to assess welfare in managed horses. Indeed, HC concentration indicates the chronic activation of the HPA axis and it is commonly employed to assess the chronic stress conditions in animal welfare studies because the concentration of cortisol in the hair reflects the history of the subject (Prandi et al., 2010). We found that NB horses had significantly lower horsehair cortisol level than TS horses, suggesting a lower activation of the HPA axis in the long term, whereas ES horses recorded intermediate values of HC and did not differ significantly from the others, which is in line with the management style that we could consider "intermediate" between TS and NB. These results are in accordance with Placci et al. (2020), who found a most favorable endocrine framework in horses kept in natural boarding, suggesting that this kind of management style matched most with the ethological and physiological needs of the species. From these results, we would have expected that NB horses also had the best judgment bias, and vice versa for the TS group, whereas NB and TS did not score differently to reach intermediate positions. Therefore, in our study horsehair cortisol did not correlate with JBT results, which is quite surprising given that JBT is considered a helpful method in the assessment of good management. Cortisol level should be discussed considering several limitations. In particular sex, breed and seasons (Aurich et al., 2015; Heimbürge et al., 2019) are the principal influencing factors that was not possible to control in this study, both for the availability of the structures and for the exclusion of most of the horses from the JBT. Analyzing also our statistical results, female and younger animal seems to have a generally lower HC level. This could have influenced the lower HC found in NB group, considering in TS group there were only geldings. Also, the season could be another potential confounding factor. Indeed, NB was tested in spring, while TS and ES in autumn and summer. Mazzola and collegues (2021) found higher cortisol concentration in the hair of horses during summer and autumn, in accordance with Aurich et al. (2015), which found that salivary cortisol level was lower in March and April. In the feces baseline, we did not find any difference between sexes and no correlation with age. Also, the season seemed to not have any influence, because baseline cortisol level did not show any difference among the 3 groups.

Conclusion

This is the first research using JBT in horses which also considers fecal and horsehair cortisol analysis. Considering the endocrinological results, with all their limitation, we cannot exclude that the procedure of the JBT itself can induce acute stress to some subjects and this is a potential confounding factor for the evaluation of the living environment on the immediate affective state of these animals. Moreover, the horsehair cortisol, which reflects the accumulated stress in the long term, did not match JBT results, thus posing even more doubts on the reliability of this test per se as applied to the evaluation of management. However, horsehair cortisol evaluation presents several limitations and potential confounding factors. According to our results, the use of the JBT test on horses seems much affected by individual motivation for food, as emerged by the high exclusion rate and by the fact that horses highly motivated for food (and present both in the natural boarding and the traditional stable management) had the best results for NP score, thus reducing the possibility to impartially evaluate the effect of management on the individuals.

In conclusion, we strongly suggest further research about the application of JBT in this species, both to ensure a higher rate of inclusion by reducing confounding factors, and to avoid the evaluation of management on its sole basis. We also recommend the selection of a multidisciplinary approach in the welfare assessment of horses.

Ethical Considerations

The study was not invasive or the animals and was evaluated and approved by the scientific Ethics Committee for Animal Welfare of DIMEVET (Department of Veterinary Medical Sciences) of the University of Bologna. The trial was monitored for Animal Welfare.

Acknowledgments

The authors thank Isabella Welsby Sjostrom for the English revision and all horses' owners and managers of the stables. This study is funded by CRUI2022.

Conflict of Interest

The authors declare no conflict of interest.

References

Accorsi, P.A., Carloni, E., Valsecchi, P., Viggiani, R., Gamberoni, M., Tamanini, C., Seren, E., 2008. Cortisol determination in hair and faeces from domestic cats and dogs. Gen. Comp. Endocrinol. 155, 398–402.

Aurich, J., Wulf, M., Ille, N., Erber, R., von Lewinski, M., Palme, R., Aurich, C., 2015. Effects of season, age, sex, and housing on salivary cortisol concentrations in horses. Dom. Anim. Endoc. 52, 11–16.

Barnard, S., Wells, D.L., Milligan, A.D.S., Arnott, G., Hepper, P.G., 2018. Personality traits affecting judgement bias task performance in dogs, (Canis familiaris). Sci. Rep. 8 (1), 6660. doi:10.1038/s41598-018-25224-y.

Beaver, B.V., 2019. Equine Behavioral Medicine. Elsevier Academic Press, London.

Bethell, E.J., 2015. A "How-To" guide for designing judgment bias studies to assess captive animal. J. App. Anim. Welf. Sci. 18 (1), S18–S42 sup.

Brooks, D.E., Komaromy, A.M., Kallberg, M.E., 1999. Comparative retinal ganglion cell and optic nerve morphology. Vet. Ophtalm. 2, 3–11.

Broom, D.M., 2003. The Evolution of Morality and Religion. Cambridge University Press, Cambridge.

Burman, O.H., Parker, R.M., Paul, E.S., Mendl, M.T., 2009. Anxiety-induced cognitive bias in non-human animals. Physiol. Behav. 98 (3), 345–350.

Christie, J.L., Hewson, C.J., Riley, C.B., McNiven, M.A., Dohoo, I.R., Bate, L.A., 2006. Management factors affecting stereotypies and body condition score in non-racing horses in Prince Edward Island. Can. Vet. J. 47, 136–143.

Daros, R.R., Costa, J.H., von Keyserlingk, M.A., Hötzel, M.J., Weary, D.M., 2014. Separation from the dam causes negative judgement bias in dairy calves. PLoS ONE 9 (5), 1–5.

Dawkins, M.S., 2006. A user’s guide to animal welfare science. Ecol. Evol. 21 (2), 77–82.

De Boyer Des Roches, A., Richard-Yris, M., Henry, S., Ezzaouïa, M., Hausberger, M., 2008. Laterality and emotions: Visual laterality in the domestic horse, Equus caballus differs with objects’ emotional value. Physiol. Behav. 94 (3), 487–490.

Duncan, I.J.H., 2006. The changing concept of animal sentience. Appl. Anim. Behav. Sci. 100, 11–19.

Duran, M.C., Janz, D.M., Waldner, C.L., Campbell, J.R., Marques, F.J., 2017. Hair cortisol concentration as a stress biomarker in horses: Associations with body location and surgical castration. J. Eq. Vet. Sci. 55, 27–33.

Freymond, S.B., Briefer, E.F., Zollinger, A., Gindrat-von Allmen, Y., Wyss, C., Bachmann, I., 2014. Behaviour of horses in a judgment bias test associated with positive or negative reinforcement. App. Anim. Behav. Sci. 158, 34–45.

Hausberger, M., Gautier, E., Biquand, V., Lunel, C., Jégo, P., 2009. Could work be a source of behavioural disorders? A study in horses. PLoS ONE 4 (10), 1–7.

Heimbürge, S., Kanitz, E., Otten, W., 2019. The use of hair cortisol for the assessment of stress in animals. Gen. Comp. Endoc. 270, 10–17.

Henneke, D.R., Potter, G.D., Kreider, J.L., Yeates, B.F., (1983 Oct). Relationship between condition score, physical measurements and body fat percentage in mares. Equine Vet J. 15 (4), 371–372. doi:10.1111/j.2042-3306.1983.tb01826.x, PMID: 6641685.

Henry, S., Fureix, C., Rowberry, R., Bateson, M., Hausberger, M., 2017. Do horses with poor welfare show ’pessimistic’ cognitive biases? Naturwiss 104 (8), 1–15.

Hintze, S., Melotti, L., Colosio, S., Bailoo, J.D., Boada-Sana, M., Würbel, H., Murphy, E., 2018. A cross-species judgement bias task: Integrating active trial initiation into a spatial Go/No-go task. Sci. Rep. 8 (1), 1–13.

Karagiannis, C.I., Burman, O.H., Mills, D.S., 2015. Dogs with separation-related problems show a “less pessimistic” cognitive bias during treatment with fuoxetine (ReconcileTM) and a behaviour modifcation plan. BMC Vet. Res. 11, 80.

Koolhaas, J.M., Bartolomucci, A., Buwalda, B., de Boer, S.F., Flugge, G., Korte, S.M., Meerlo, P., Murison, R., Olivier, B., Palanza, P., Richter-Levin, G., Sgoifo, A., Steimer, T., Stiedl, O., van Dijk, G., Wohr, M., Fuchs, E., 2011. Stress revisited: A critical evaluation of the stress concept. Neur. Biobehav. Rev. 35, 1291–1301.

Lalot, M., Ung, D., Péron, F., d’Ettorre, P., Bovet, D., 2017. You know what? I’m happy Cognitive bias is not related to personality but is induced by pair-housing in canaries, Serinuscanaria. Behav. Process 134, 70–77.

Lansade, L., Bouissou, M.-F., 2008. Reactivity to humans: A temperament trait of horses which is stable across time and situations. App. Anim. Behav. Sci. 114, 492–508.

Lloyd, A.S., Martin, J.E., Imogen-Bornett-Gauci, H.L., Wilkinson, R.G., 2007. Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. App. Anim. Behav. Sci. 105 (1), 205–222 –3.

Löckener, S., Reese, S., Erhard, M., Wöhr, A., 2016. Pasturing in herds after housing in horseboxes induces a positive cognitive bias in horses. J. Vet. Behav. 11, 50–55.

Marliani, G., Sprocatti, I., Schiavoni, G., Bellodi, A., Accorsi, P.A., 2021. Evaluation of horses' daytime activity budget in a model of ethological stable: A case study in Italy. J. App. Anim. Welf. Sci. doi:10.1080/10888705.2020.1857252.

Mal, M.E., Friend, T.H., Lay, D.C., Vogelsang, S.G., Jenkins, O.C., 1991. Behavioural responses of mares to short term confinement and social isolation. Appl. Anim. Behav. Sci. 31, 13–24.

Mazzola, S.M., Colombani, C., Pizzamiglio, G., Cannas, S., Palestrini, C., Costa, E.D., Gazzonis, A.L., Bionda, A., Crepaldi, P., 2021. Do you think I am Llving well? A four-season hair cortisol analysis on leisure horses in different housing and management conditions. Animals. 11, 2141.

McCall, C.A., 1989. The effect of body condition of horses on discrimination learning abilities. Appl. Anim. Behav. Sci. 22 (3-4), 327–334.

McGreevy, P., 2012. Equine Behavior, 2nd edition Elsevier, China.

McIlwraith, W.C., Rollin, B.E., 2011. Equine Welfare. Blackwell Publishing, Hoboken, New Jersey.

Mellor, D.J., Beausoleil, N.J., 2015. Extending the "Five Domains" model for animal welfare assessment to incorporate positive welfare states. Anim. Welf. 24, 241–253.

Mellor, D.J., Beausoleil, N.J., Littlewood, K.E., McLean, A.N., McGreevy, P.D., Jones, B., Wilkins, C., 2020. The 2020 Five Domains Model: Including human–animali nteractions in assessments of animal welfare. Anim 10, 1870.

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., Casey, R., 2010. Dogs showing separation-related behaviour exhibit a ‘pessimistic’ cognitive bias. Curr. Biol. 20, 839–840.

Mendl, M., Burman, O.H.P., Parker, R.M.A., Paul, E.S., 2009. Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. App. Anim. Behav. Sci. 118 (3), 161–181 –4.

Mostl, E., Messmann, S., Bagu, E., Robia, C., Palme, R., 1999. Measurement of glucocorticoid metabolite concentrations in faeces of domestic livestock. Zentral. A. 46 (10), 621–631.

Motta, M., Degli Esposti, A., 1981. A computer program for mathematical treatment of data in radioimmunoassay. Comput. Progr. Biomed. 13 (1), 121–129 -2.

Neave, H.W., Daros, R.R., Costa, J.H., Von Keyserlingk, M.A., Weary, D.M., 2013. Pain and pessimism: Dairy calves exhibit negative judgement bias following hot-iron disbudding. PLoS One 9 (4), 1–6.

Palme, R., 2012. Monitoring stress hormone metabolites as a useful, non-invasive tool for welfare assessment in farm animal. Anim. Welf. 21, 331–337.

Paul, E.S., Harding, E.J., Mendl, M., 2005. Measuring emotional processes in animals: The utility of a cognitive approach. Neurosci. Biobehav. Rev. 29, 469–491.

Placci, M., Marliani, G., Sabioni, S., Gabai, G., Mondo, E., Borghetti, P., De Angelis, E., Accorsi, P.A., 2020. Natural horse boarding vs traditional stable: A comparison of hormonal, hematological and immunological parameters. J. App. Anim. Welf. Sci. 23 (3), 366–377.

Prandi, A., Comin, A., Peric, T., Montillo, M., Omodeo, S.G., 2010. Nuovo approccio non invasivo per la valutazione del benessere animale. Quaderno SOZOOALP 6, 183–191.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ralph, C.R., Tilbrook, A.J., 2016. Invited review: The usefulness of measuring glucocorticoids for assessing animal welfare. J. Anim. Sci. 94, 457–470.

Roelofs, S., Boleij, H., Nordquist, R.E., van der Staay, F.J., 2016. Making decisions under ambiguity: Judgment bias tasks for assessing emotional state in animals. Front Behave. Neurosci 10 (119), 1–16.

Ruet, A., Lemarchand, J., Parias, C., Mach, N., Moisan, M., Foury, A., Briant, C., Lansade, L., 2019. Housing horses in individual boxes is a challenge with regard to welfare. Anim 9 (9), 621–640.

Saba, C., Ascione, L., Montignani, A., 2013. Guida Pratica Alla Gestione Naturale Del Cavallo. EquiLuna asd, Fivizzano, MS.

Sheriff, M.J., Dantzer, B., Delehanty, B., Palme, R., Boonstra, R., 2011. Measuring stress in wildlife: Techniques for quantifying glucocorticoids. Oecologia. 166, 869–887.

Tamanini, C., Giordano, N., Chiesa, F., Seren, E., 1983. Plasma cortisol variations induced in the stallion by mating. Acta. End 102 (3), 447–450.

Vallortigara, G., Rogers, L.J., 2005. Survival with an asymmetrical brain: Advantages and disadvantages of cerebral lateralization. Behav. Brain Sci. 28, 575–589.

Whittaker, A.L., MarshBarker, T.L.H.E., 2020. A consideration of the role of biology and test design as confounding factors in judgement bias tests: The role of behavioural assessment in determining 'positive' affective states in animals. CAB Rev. App. Anim. Behav. Sci. 23214, 1051261 -13.

Wolff, A., Hausberger, M., Le Scolan, N., 1997. Experimental tests to assess emotionality in horses. Behav. Proc. 40, 209–221.

Yarnell, K., Hall, C., Royle, C., Walker, S.L., 2015. Domesticated horses differ in their behavioural and physiological responses to isolated and group housing. Physiol. Behav. 143, 51–57.

Yngvesson, J., Rey Torres, J., Lindholm, J., Pättiniemi, A., Andersson, P., Sassner, H., 2019. Health and body conditions of riding school horses housed in groups or kept in conventional tie-stall/box housing. Anim 9 (3), 73.

Zoller, J.L., Cavinder, C.A., Sigler, D., Tedeschi, L.O., Harlin, J., 2019. Development of a mathematical model for predicting digestible energy intake to meet desired body condition parameters in exercising horses. J. Anim. Sci. 97, 1945–1955.

Tables

Table 1. The number of total mares, geldings, and stallions for each stable and mean age ± standard deviation of each group considered for the study. In brackets the number of mares, gelding and stallions and mean age ± standard deviation of horses that completed the test.

	Natural Boarding	**Traditional Stable**	**Ethological Stable**
Mares	7 (4)	4 (0)	5 (1)
Geldings	9 (6)	9 (6)	6 (3)
Stallions	-	1 (0)	-
Age (years)	14 ± 6 (12± 6)	18 ± 5 (17± 4)	18 ± 6 (19± 3)

Table 2. Results of the Kruskal-Wallis test with Dunn's adjustments for multiple comparisons of fecal cortisol concentration (pg/mg) recorded in each phase of the test, considering a time-lag of 24 hours from their collection. The median ± IQR were reported for each kind of management, and the letters indicate the different groups according to Dunn's test.

Phase	**Ethological Stable**	**Natural Boarding**	**Traditional Stable**
P1 (baseline)	3.04 ±1.57^{a}	1.71±4.40 ^{a}	2.33±0.47 ^{a}
P2 (habituation)	3.06±0.42 ^{b}	6.80±1.86 ^{a}	3.26±3.39 ^{b}
P3 (training)	5.92±2.63 ^{a}	7.52±5.21 ^{a}	3.49±1.34 ^{b}
P4 (test1)	7.08±4.22 ^{a,b}	9.71±3.12 ^{a}	2.45±1.74 ^{b}

a,b = $p<0.05$

Figures

Figure 1. Representation of the test arena. P = Positive location; NP = Near Positive intermediate location; M = Medium intermediate location; NN = Near Negative intermediate location; N = Negative location. P and N were randomly on the left side or the right side of the horse.

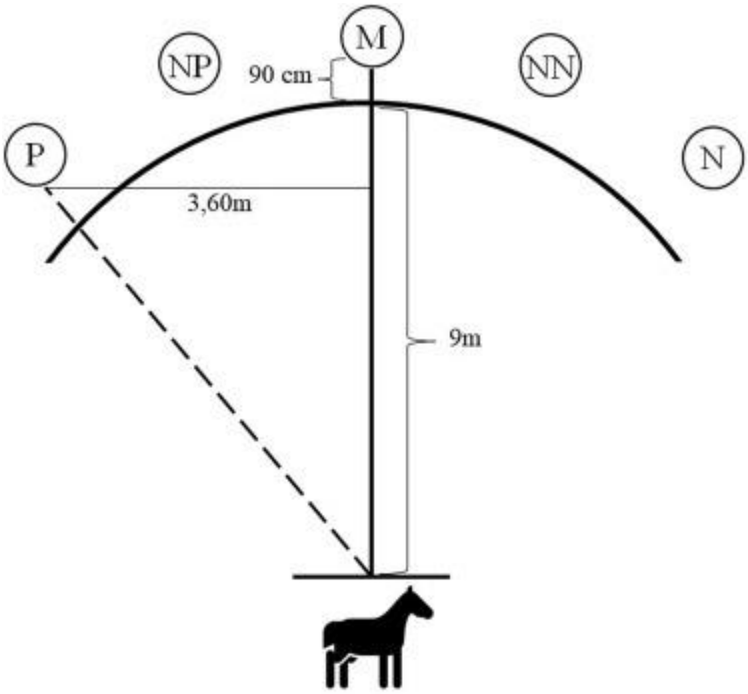

Figure 2. Boxplot of the Body Condition Score (BCS) of excluded and included horses. The box plots show the median and 25th and 75th percentiles; the whiskers indicate the values within 1.5 times the inter-quartile range, IQR. The outlier is indicated by the black circle; *P*-value < 0.05 is indicated by *.

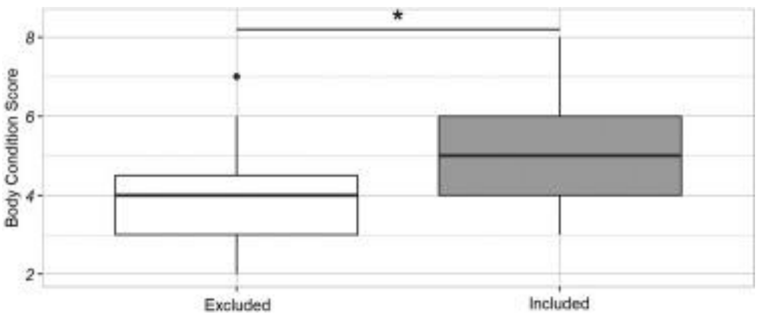

Figure 3. Boxplot of JBT results of each group (ES = ethological stable; NB = natural boarding system; TS = traditional stable). The box plots show the median and 25th and 75th percentiles; the whiskers indicate the values within 1.5 times the inter-quartile range, IQR. The outlier is indicated by the black circle. P-value = 0.01is indicated by **, and P-value < 0.05 is indicated by *.