



## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

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

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Is the judgment bias test a good tool to assess the quality of horse management? / Marliani, G; Balboni, A; Tiberi, C; Malavasi, R; Gardini, A; Accorsi, PA. - In: JOURNAL OF VETERINARY BEHAVIOR. - ISSN 1558-7878. - STAMPA. - 58:December(2022), pp. 62-69. [10.1016/j.jveb.2022.11.002]

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>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

(Article begins on next page)

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

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

2 Giovanna Marliani\*<sup>a</sup>, Alessia Balboni<sup>a</sup>, Carlotta Tiberi<sup>a</sup>, Rachele Malavasi<sup>b</sup>, Aldo Gardini<sup>c</sup>,  
3 Pier Attilio Accorsi<sup>a</sup>

4 <sup>a</sup>Department of Veterinary Medical Sciences, University of Bologna, Ozzano dell'Emilia  
5 (BO), Italy

6 <sup>b</sup>School of Ethical Equitation, Moncigoli di Fivizzano, +3498551997, Italy

7 <sup>c</sup>Department of Statistical Sciences P. Fortunati, University of Bologna, Bologna, Italy

## 8 **Abstract**

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

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

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

## 37 **Introduction**

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

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

55 However, stress and, consequently, cortisol levels can be influenced by both emotional and  
56 physical stressors. Therefore, to investigate emotions in non-vocal animals, research relies on  
57 indirect indications of how the animal perceives the environment, considering that emotional  
58 states can influence memory, attention, and judgment (Duncan, 2006; Mendl et al., 2009). In  
59 particular, Judgment bias tests (JBTs), modulated by human sciences (Bethell, 2015), have  
60 been employed in animal research to investigate how the environment or routine procedures

61 affect animals' affective states (Burman et al., 2009, Neave et al., 2013; Daros et al., 2014).  
62 In animal research, judgment bias represents the animal's response to a neutral stimulus and,  
63 evaluating if the individual expects from this stimulus either a positive or a negative outcome,  
64 it defines the subject as optimistic or pessimistic (Mendl et al., 2009).

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

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

## 91 **Material and Methods**

92 *Stables*

93 Natural Boarding System (NB)

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

104 Traditional Stable (TS)

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

116 Ethological Stable (ES)

117 The system was the same as described by Marliani et al., 2021. During the study, the stable  
118 hosted 11 horses. The stable included 12 stalls, called Big Box© (8.00 m x 5.00 m), arranged  
119 in rows of 6. Each stall was divided into 2 sections of 20m<sup>2</sup>: the front part consisted of high-  
120 density impact rubber mats that provided the horse with rigid support; the rear part consisted  
121 of a soft and semi-permanent element of river sand. This system of double litter allowed the

122 horse freedom of movement and choice of the ground on which to rest. The walls of the Big  
123 Box© were of metal gratings, with windows between stalls, so that subjects were able to see  
124 and touch each other. There was a paddock of 1.50ha organized with large areas and  
125 corridors, and 6 paddocks of 2.00ha each. Horses, individually or in a small group, stayed in  
126 paddocks all day and were brought into the stall for the evening, or in case of bad weather.  
127 Hay was available throughout the day and cereals (barley, oats, and corn), apples, and carrots  
128 were provided at 7.00 a.m. and 7.00 p.m., according to a personalized food plan. There were  
129 also a round sand pen and a covered sand rectangular field. Horses were engaged principally  
130 in flatwork, country riding, and walks, and they were not used for competitions.

### 131 *Subjects*

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

### 136 *Body Condition Score (BCS) and Personality Evaluation*

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

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

153 *Judgment Bias Test*

154 The protocol employed in the study was a modification of those proposed by Freymond et al.  
155 (2014) and Henry et al. (2017). It was a Go/No-Go JBT with spatial stimuli. Black buckets  
156 were employed for the test, to prevent the horses seeing the contents. In each stable, the test  
157 was conducted in a fenced arena with a sandy substrate, familiar to the subjects and  
158 sufficiently far from disturbances.

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

174 *Habituation*

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

182 *Training*

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

### 195 *Testing*

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

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

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

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



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

213 
$$\text{Adjusted latency} = (\text{mean latency to ambiguous location} - \text{mean latency to P}) / (\text{mean latency}$$
  
214 
$$\text{to N} - \text{mean latency to P})$$

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

### 216 *Collection of Fecal and Horsehair Samples*

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

- 219 1. the habituation (P1), which was representative of the baseline cortisol concentration before  
220 the beginning of the whole procedure;
- 221 2. the training (P2), which was representative of cortisol level during habituation;
- 222 3. the first session of the test (P3), which was representative of training;
- 223 4. the second session of the test (P4), representative of the first day of the test.

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

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

### 234 *Cortisol Assay*

235 Cortisol concentrations were determined by Radio Immuno-assays (RIAs). Cortisol was  
236 extracted from fecal specimens (500 mg, wet weight) with methanol-water solution (v/v 4:1)

237 and ethyl ether, whereas for horsehairs 60 mg of trimmed horsehairs (1-3 mm) of a subject  
238 were put in a glass vial with 5 ml methanol (Accorsi et al., 2008). All samples were dried  
239 under an air-stream suction hood at 37°C and the dry-residue was dissolved into phosphate-  
240 buffered saline (PBS) 0.05 M, pH 7.5. Cortisol metabolites assay in both feces and horsehair  
241 were carried out according to Tamanini et al. (1983). The cortisol RIA was performed using  
242 an antiserum to cortisol-21-hemisuccinate-BSA (anti-rabbit), at a working dilution of 1:20  
243 000 and 3H-cortisol (30 pg/tube vial) as a tracer. Validation parameters of analysis were:  
244 sensitivity 0.19 pg/mg, intra-assay variability 5.9%, inter-assay variability 8.7%.  
245 Radioactivity was determined using a liquid scintillation  $\beta$  counter and a linear standard  
246 curve, ad hoc designed by a software program (Motta and Degli Esposti, 1981).

#### 247 *Statistical Analysis*

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

## 259 **Results**

### 260 *Task Acquisition*

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

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

269 Twenty-two horses had the positive bucket on their left side, 19 on their right. Considering  
270 this variable and the entire group in a chi-squared test, there was a significantly higher  
271 probability that the horses who had the P location on their left side was excluded from the test  
272 ( $\chi^2 = 5.47$ ,  $df = 1$ ,  $P\text{-value} = 0.02$ ). Indeed, considering the subjects included in the test, 13  
273 had P on their right side and 7 on their left.

274 The BCS of the excluded horses was significantly ( $W = 104$ ;  $r = -0.359$ ; 95% C.I. [-0.614, -  
275 0.109];  $P\text{-value} = 0.02$ ) lower (Median  $\pm$  IQR:  $4 \pm 1.5$ ,  $n = 19$ ) than that of the included  
276 subjects ( $5 \pm 2.0$ ,  $n = 19$ ) (Figure 2).

### 277 *JBT Scores*

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

284 Considering the entire group of horses, the scores recorded in NP (Median  $\pm$  IQR:  $10.24 \pm$   
285  $44.63$ ), M ( $2.88 \pm 17.34$ ), and NN locations ( $71.09 \pm 67.00$ ) were significantly different,  
286 according to the Kruskal-Wallis test with Dunn's post-hoc test ( $\chi^2 = 18.7236$ ,  $df = 2$ ,  $P\text{-value} <$   
287  $0.001$ ). In particular, NN was significantly different from NP ( $P\text{-value} < 0.01$ ) and M ( $P\text{-}$   
288  $\text{value} < 0.001$ ), but NP and M did not differ ( $P\text{-value} = 0.34$ ). No correlation was found  
289 between the scores recorded for the intermediate positions and personality traits and age, but  
290 there was a significant negative correlation between the score in NP and BCS ( $S = 1907.1$ ,  $P$   
291  $< 0.01$ ,  $\rho = -0.67$ ). Body condition score differs between groups ( $\chi^2 = 9.0149$ ,  $df = 2$ ,  $P\text{-}$   
292  $\text{value} = 0.01$ ) and it was significantly higher in NB ( $6.22 \pm 1.72$ ) horses than in TS ones ( $3.83$   
293  $\pm 0.75$ ,  $P\text{-value} = 0.01$ ), while ES subjects recorded an intermediate value ( $5 \pm 0$ ), and did not  
294 differ neither from TS ( $P\text{-value} = 0.14$ ) and NB ( $P\text{-value} = 0.64$ ).

### 295 **Cortisol Results**

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

301 The FC baseline concentration was not significantly different between the 3 groups of horses  
302 ( $\chi^2=0.56$ ,  $df=2$ ,  $P\text{-value}=0.75$ ). However, during habituation ( $\chi^2=6.67$ ,  $df=2$ ,  $P\text{-}$   
303  $\text{value}=0.04$ ), training ( $\chi^2=11.59$ ,  $df=2$ ,  $P\text{-value}<0.001$ ) and the first day of the test ( $\chi^2=$   
304  $11.97$ ,  $df=2$ ,  $P\text{-value}<0.001$ ), subjects from NB registered a significantly higher  
305 concentration of FC than the TS group. The ES group showed a significantly higher  
306 concentration than TS horses in P3 ( $P\text{-value}=0.03$ ) and, with a significant tendency, in P4  
307 ( $P\text{-value}=0.07$ ). (Table 2).

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

## 320 **Discussion**

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

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

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

360 are used for. Indeed, in the 3 management we noticed that NB horses had a significant higher  
361 BCS than TS. Not only feeding routine, health status and environment can influence BCS,  
362 but also the type of exercise (Christie et al., 2006; Zoller et al., 2019). Therefore, the  
363 differences that we found in the 3 systems can be also due to the differences of equestrian  
364 activities in which horses were involved.

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

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

386 Secondly, we predicted that different management styles would have affected JBT results.  
387 Only the ES group showed a significantly lower latency to reach NN, suggesting that those  
388 horses had greater anticipation of a positive outcome at that location, and thus experienced a  
389 general more positive affective state. NN is usually associated to negative outcome and a  
390 slower latency to reach this probe location is usually interpreted as index of fear and anxiety,  
391 so of a pessimistic affective state (Barnard et al., 2018). Indeed, our results are similar to

392 those of Karagiannis and colleagues (2015) that found that dogs suffered from separation  
393 related problems approach significantly slower NN probes but not the other locations.  
394 However, this result should be considered with caution, because only 4 of 11 horses  
395 completed the JBT in the ethological stable. Contrary to our predictions, NB and TS did not  
396 score differently to reach intermediate positions, although these 2 management systems are  
397 extremely different. It cannot be excluded that the manipulation by strangers of NB horses  
398 and, overall, the separation from their herd caused a mild state of distress, which could have  
399 altered the results of JBT. This hypothesis can be partially confirmed by the analysis of fecal  
400 cortisol, which evidenced an increasing level of glucocorticoids in the feces (representative of  
401 daily levels of stress) and a significantly higher FC concentration in NB in respect to TS  
402 subjects during the JBT procedure. According to Roelofs et al. (2016), involving social  
403 animals individually in JBT could increase stress perception. Indeed, it is well known that  
404 social isolation induces stress in most gregarious species, manifested through behavioral and  
405 physiological changes such as increasing vocalization, locomotion, heart rate, and plasma  
406 cortisol levels (Mal et al., 1991; Wolff, Hausberger and Le Scolan, 1997; Lansade and  
407 Bouissou, 2008), and the horse is a social animal that forms preferential bonds within the  
408 herd to guarantee social stability (Beaver, 2019). Therefore, the procedure itself could have  
409 been stressful for NB horses, which are also usually less used to being manipulated by  
410 strangers, and this can have determined that they did not differ from the TS group in their  
411 latency to reach intermediate positions. All these elements suggest that JBT may not be  
412 reliable to evaluate the quality of the management style for these individuals because it is  
413 affected by contingent confounding factors.

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

424 cortisol allowed us to monitor the stress of animals possibly caused by the entire procedure of  
425 JBT, considering also that it could be a stressful disruption of their daily routine.

426 We obtained surprising results when we considered horsehair cortisol concentration, whose  
427 results, if matching with those of JBT, would have reinforced the role of JBT as a useful  
428 method to assess welfare in managed horses. Indeed, HC concentration indicates the chronic  
429 activation of the HPA axis and it is commonly employed to assess the chronic stress  
430 conditions in animal welfare studies because the concentration of cortisol in the hair reflects  
431 the history of the subject (Prandi et al., 2010). We found that NB horses had significantly  
432 lower horsehair cortisol level than TS horses, suggesting a lower activation of the HPA axis  
433 in the long term, whereas ES horses recorded intermediate values of HC and did not differ  
434 significantly from the others, which is in line with the management style that we could  
435 consider “intermediate” between TS and NB. These results are in accordance with Placci et  
436 al. (2020), who found a most favorable endocrine framework in horses kept in natural  
437 boarding, suggesting that this kind of management style matched most with the ethological  
438 and physiological needs of the species. From these results, we would have expected that NB  
439 horses also had the best judgment bias, and vice versa for the TS group, whereas NB and TS  
440 did not score differently to reach intermediate positions. Therefore, in our study horsehair  
441 cortisol did not correlate with JBT results, which is quite surprising given that JBT is  
442 considered a helpful method in the assessment of good management. Cortisol level should be  
443 discussed considering several limitations. In particular sex, breed and seasons (Aurich et al.,  
444 2015; Heimbürge et al., 2019) are the principal influencing factors that was not possible to  
445 control in this study, both for the availability of the structures and for the exclusion of most of  
446 the horses from the JBT. Analyzing also our statistical results, female and younger animal  
447 seems to have a generally lower HC level. This could have influenced the lower HC found in  
448 NB group, considering in TS group there were only geldings. Also, the season could be  
449 another potential confounding factor. Indeed, NB was tested in spring, while TS and ES in  
450 autumn and summer. Mazzola and colleagues (2021) found higher cortisol concentration in the  
451 hair of horses during summer and autumn, in accordance with Aurich et al. (2015), which  
452 found that salivary cortisol level was lower in March and April. In the feces baseline, we did  
453 not find any difference between sexes and no correlation with age. Also, the season seemed  
454 to not have any influence, because baseline cortisol level did not show any difference among  
455 the 3 groups.



## 456 **Conclusion**

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

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

## 474 **Ethical Considerations**

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

## 478 **Acknowledgments**

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

## 481 **Conflict of Interest**

482 The authors declare no conflict of interest.

## 483 **References**

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

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

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

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

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

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

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

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

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

505 Daros, R.R., Costa, J.H., von Keyserlingk, M.A., Hötzel, M.J., Weary, D.M., 2014.  
506 Separation from the dam causes negative judgement bias in dairy calves. *PLoS ONE* 9 (5), 1–  
507 5.

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

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

512 Duncan, I.J.H., 2006. The changing concept of animal sentience. *Appl. Anim. Behav. Sci.*  
513 100, 11–19.

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

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

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

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

524 Henneke, D.R., Potter, G.D., Kreider, J.L., Yeates, B.F., (1983 Oct). Relationship between  
525 condition score, physical measurements and body fat percentage in mares. *Equine Vet J.* 15  
526 (4), 371–372. doi:10.1111/j.2042-3306.1983.tb01826.x, PMID: 6641685.

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

529 Hintze, S., Melotti, L., Colosio, S., Bailoo, J.D., Boada-Sana, M., Würbel, H., Murphy, E.,  
530 2018. A cross-species judgement bias task: Integrating active trial initiation into a spatial  
531 Go/No-go task. *Sci. Rep.* 8 (1), 1–13.

532 Karagiannis, C.I., Burman, O.H., Mills, D.S., 2015. Dogs with separation-related problems  
533 show a “less pessimistic” cognitive bias during treatment with fluoxetine (Reconcile™) and  
534 a behaviour modification plan. *BMC Vet. Res.* 11, 80.

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

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

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

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

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

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

552 Mal, M.E., Friend, T.H., Lay, D.C., Vogelsang, S.G., Jenkins, O.C., 1991. Behavioural  
553 responses of mares to short term confinement and social isolation. *Appl. Anim. Behav. Sci.*  
554 31, 13–24.

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

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

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

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

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

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

569 Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., Casey, R., 2010. Dogs  
570 showing separation-related behaviour exhibit a ‘pessimistic’ cognitive bias. *Curr. Biol.* 20,  
571 839–840.

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

575 Mostl, E., Messmann, S., Bagu, E., Robia, C., Palme, R., 1999. Measurement of  
576 glucocorticoid metabolite concentrations in faeces of domestic livestock. *Zentral. A.* 46 (10),  
577 621–631.

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

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

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

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

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

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

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

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

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

599 Ruet, A., Lemarchand, J., Parias, C., Mach, N., Moisan, M., Foury, A., Briant, C., Lansade,  
600 L., 2019. Housing horses in individual boxes is a challenge with regard to welfare. *Anim* 9  
601 (9), 621–640.

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

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

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

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

610 Whittaker, A.L., MarshBarker, T.L.H.E., 2020. A consideration of the role of biology and test  
611 design as confounding factors in judgement bias tests: The role of behavioural assessment in  
612 determining 'positive' affective states in animals. *CAB Rev. App. Anim. Behav. Sci.* 23214,  
613 1051261 -13.

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

616 Yarnell, K., Hall, C., Royle, C., Walker, S.L., 2015. Domesticated horses differ in their  
617 behavioural and physiological responses to isolated and group housing. *Physiol. Behav.* 143,  
618 51–57.

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

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

625 **Tables**

626 **Table 1.** The number of total mares, geldings, and stallions for each stable and mean age ±  
 627 standard deviation of each group considered for the study. In brackets the number of mares,  
 628 gelding and stallions and mean age ± standard deviation of horses that completed the test.  
 629

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

630  
 631 **Table 2.** Results of the Kruskal-Wallis test with Dunn's adjustments for multiple comparisons  
 632 of fecal cortisol concentration (pg/mg) recorded in each phase of the test, considering a time-  
 633 lag of 24 hours from their collection. The median ± IQR were reported for each kind of  
 634 management, and the letters indicate the different groups according to Dunn's test.

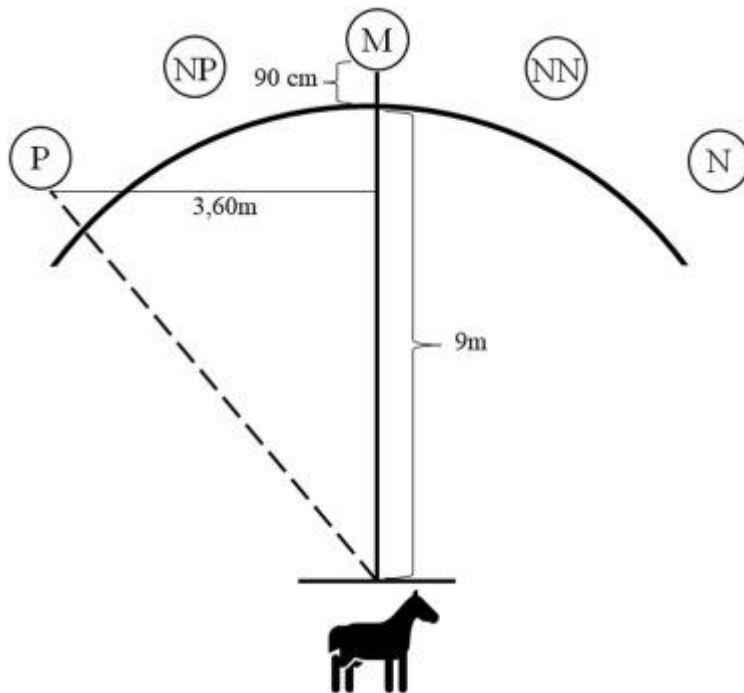
<b>Phase</b>	<b>Ethological Stable</b>	<b>Natural Boarding</b>	<b>Traditional Stable</b>
P1 (baseline)	3.04 ±1.57 <sup>a</sup>	1.71±4.40 <sup>a</sup>	2.33±0.47 <sup>a</sup>
P2 (habituation)	3.06±0.42 <sup>b</sup>	6.80±1.86 <sup>a</sup>	3.26±3.39 <sup>b</sup>
P3 (training)	5.92±2.63 <sup>a</sup>	7.52±5.21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

635 a,b = p<0.05

636

637 **Figures**

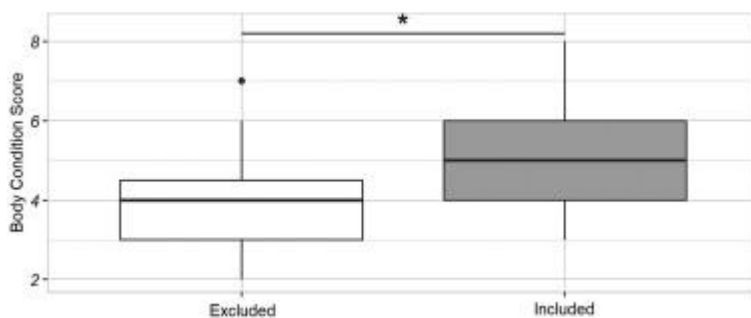
638 **Figure 1.** Representation of the test arena. P = Positive location; NP = Near Positive  
639 intermediate location; M = Medium intermediate location; NN = Near Negative intermediate  
640 location; N = Negative location. P and N were randomly on the left side or the right side of  
641 the horse.



642

643

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



648

649

650

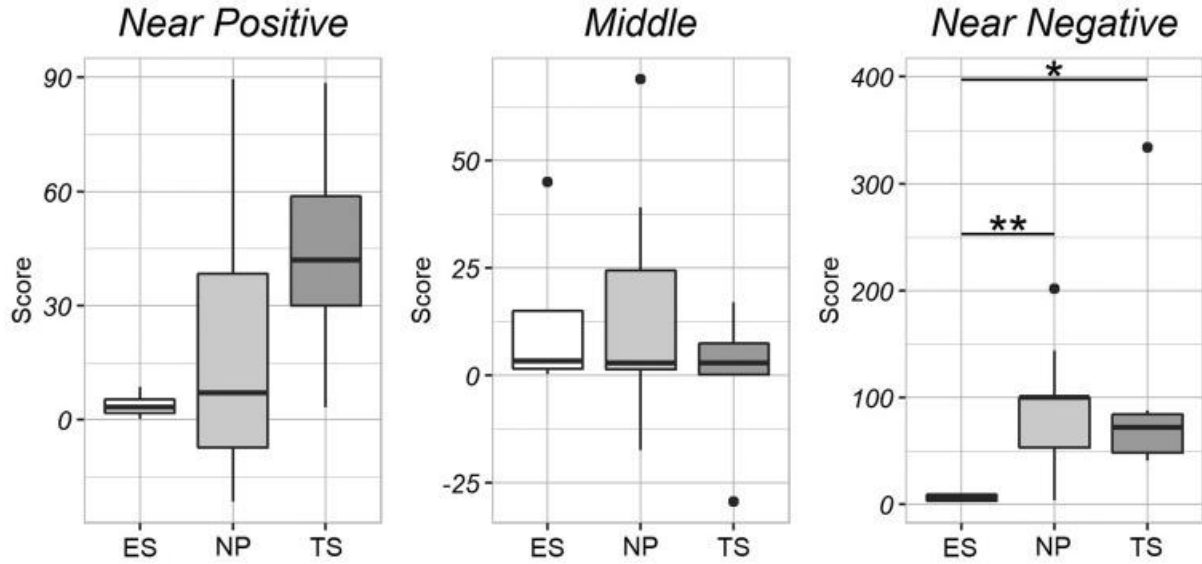
651

652

653



654 **Figure 3.** Boxplot of JBT results of each group (ES = ethological stable; NB = natural  
 655 boarding system; TS = traditional stable). The box plots show the median and 25th and 75th  
 656 percentiles; the whiskers indicate the values within 1.5 times the inter-quartile range, IQR.  
 657 The outlier is indicated by the black circle. P-value = 0.01 is indicated by \*\*, and P-value <  
 658 0.05 is indicated by \*.



659