9-2014

Behaviour of Horses in a Judgment Bias Test Associated with Positive or Negative Reinforcement

Sabrina Briefer Freymond
Agroscope - Swiss National Stud Farm

Elodie F. Briefer
Queen Mary University of London

Anja Zollinger
Bern University of Applied Sciences

Yveline Gindrat-von Allmen
Bern University of Applied Sciences

Christa Wyss
Agroscope - Swiss National Stud Farm

See next page for additional authors

Authors
Sabrina Briefer Freymond, Elodie F. Briefer, Anja Zollinger, Yveline Gindrat-von Allmen, Christa Wyss, and Iris Bachmann

Behaviour of Horses in a Judgment Bias Test Associated with Positive or Negative Reinforcement

Sabrina Briefer Freymond[1], Elodie F. Briefer[2], Anja Zollinger[1,3], Yveline Gindrat-von Allmen[3], Christa Wyss[1], Iris Bachmann[1]

[1] *Agroscope – Swiss National Stud Farm*
[2] *ETH Zürich*
[3] *Bern University of Applied Sciences*

KEYWORDS

cognitive bias, *Equus caballus*, emotions, optimism, training method

ABSTRACT

Moods can influence our judgment of ambiguous stimuli as positive or negative. Measuring judgment bias in animals is a promising method to objectively assess their emotional states. Our study aimed to develop a cognitive bias test in horses, in order to assess the effect of training using positive reinforcement (PR) or negative reinforcement (NR) on their emotional states. We trained 12 mares to discriminate between a rewarded and a non-rewarded location situated on each side of a paddock. The mares were then trained during five days to perform several exercises using PR (n = 6) for one group, and NR (n = 6) for the other (treatment). Finally, we compared the responses of the two groups to three ambiguous locations situated between the rewarded and non-rewarded locations (judgment bias test). During the training exercises, according to our predictions, behavioural measures suggested that NR mares experienced more negative emotions than PR mares. Surprisingly, the results of the judgment bias test suggest that NR mares were in a more optimistic mood compared to PR mares, despite previously experiencing more negative emotions during the treatment. NR mares could have been more motivated to obtain a food reward than PR mares, which had been rewarded throughout the treatment phase. Alternatively, NR mares could have developed optimistic bias triggered by release from the negative state experienced during treatment. This first attempt to test judgment bias in horses suggests that this is a promising method to measure horse mood. Knowledge about the effect of training methods on the mental health of domesticated animals can add a new dimension to animal welfare, in order to promote better ways to work with animals.

1. Introduction

Public concern about animal welfare is strongly based on the attribution of mental states to animals (Mendl and Paul, 2004), and animal welfare is nowadays assessed via both physical and mental health (Dawkins, 2008). Negative emotional states, and particularly mood disorders, should be minimized. Good welfare thus implies a reduction in animal suffering and other negative emotions (e.g. fear; Forkman et al., 2007). In recent years, the importance of promoting positive emotions has also been suggested (Boissy et al., 2007). However, animal welfare legislation in many countries focusses only on how animals are housed and not on how people interact with them (e.g. during training). Few studies have investigated the impact of human interactions on animal mental health (Hausberger et al., 2009; Sankey et al., 2010a, 2010b).

Emotions are intense but short-lived affective reactions to events, and have an important function for animals' lives. They act as "detectors" to identify situations and select appropriate behavioural decisions (Mendl et al., 2010a, Nettle and Bateson, 2012). Emotions can be defined using two dimensions: their arousal (bodily activation or excitation) and their valence (positive or negative), and four components: cognitive, behavioural, neurophysiological and subjective component (Keltner and Lerner, 2010; Scherer, 1984). The subjective component of emotions cannot yet be shown to occur in animals. However, cognitive, behavioural and neurophysiological components are increasingly used as indicators of animal emotions (Mendl et al., 2010a). The behavioural parameters that can be affected by emotions (among other factors) include body, tail and ear postures, movements and also vocalizations (e.g. Boissy et al., 2011; Briefer, 2012; Reefmann et al., 2009; Young et al., 2012). Many studies have revealed indicators of emotional arousal, and these largely focused on arousal during situations of negative rather than positive valence. In contrast, the evidence suggests that it might be difficult to find clear indicators of valence (Mendl et al., 2010a). Yet, indicators of valence would be especially useful to differentiate between negative (e.g. behavioural despair) and positive (e.g. relaxed animal) states of low arousal, which are characterized by quite similar physiological and behavioural expressions (e.g. lack of movement, low head posture; Fureix et al., 2012; Hall et al., 2008). Therefore, more research is needed to highlight indicators of emotional states, and particularly of emotional valence, in order to assess animal welfare.

One promising method to differentiate between positive and negative emotional states is the cognitive bias approach (Désiré et al., 2002; Mendl et al., 2009; Paul et al., 2005). This method assesses the valence of long-term emotional states (i.e. mood) through their impact on cognitive processes (i.e. attention, learning, memory and decision-making). Moods are diffuse emotional states that arise as a result of an accumulation of shorter term emotional states (Mendl et al., 2010a, Nettle and Bateson, 2012). In humans, individuals in positive moods will be biased towards positive events and be more attentive to them, learn them more easily, memorize them better and expect them more often than negative events ("optimistic bias"), whereas the opposite happens for people in negative moods ("pessimistic bias"; MacLeod and Byrne, 1996; Sharot, 2011; Strunk et al., 2006; Wright and Bower, 1992). In non-human animals, "pessimism" and "optimism" can be assessed by examining how animals perceive "ambiguous" stimuli that are intermediate between two "reference" stimuli, such as a stimulus known to be rewarded and another one known to be unrewarded or aversive ("judgment bias"; Harding et al., 2004; Mendl et al., 2009). This judgment bias technique has now been successfully tested in several species, from honeybees (*Apis mellifera carnica*; Bateson et al., 2011), to European starlings (*Sturnus vulgaris*; Brilot et al., 2010), dogs (*Canis lupus* Mendl et al., 2010b) and goats (*Capra hircus*; Briefer and McElligott, 2013; see also review Mendl et al., 2009).

The mental health of domestic animals is likely not only affected by their housing conditions (Bateson and Matheson, 2007; Brydges et al., 2011; Douglas et al., 2012; Matheson et al., 2008), but also by the way that people behave towards them. Specifically, the way animals are trained (e.g. shepherding in dogs, riding in donkeys (*Equus asinus*), horses (*Equus caballus*) or camels (*Camelus spp*.)) could influence their mood (Hall et al., 2008; Hausberger et al., 2009). This may especially be true for horses, which are trained for various tasks (e.g. jumping, dressage) on a regular basis. For instance, controversial or unclear commands from the rider can lead to frustration, neurosis and even chronic stress in horses, resulting in an apathetic state or stereotypies (Hausberger et al., 2009).

Animals can be trained using positive (PR) or negative reinforcement (NR). PR consists of rewarding the animal, typically with food, when it performs the requested behaviour. Conversely, NR consists of applying an aversive stimulus and ending it when the requested behaviour is achieved (Skinner, 1938). NR is often mistaken for punishment and, if not carefully applied, this training method can trigger fear (Waran et al., 2002). This method is more commonly used in animal training than PR (Innes and McBride, 2008). Recent evidence suggests that training using PR triggers a more positive perception of humans than NR (Sankey et al., 2010a). When PR is used, the animal learns a requested behaviour through motivation to obtain and consume a reward. This type of training thus involves appetitive-anticipation, consumption and post-consumption of a reward, which can induce a positive emotional state (e.g. satisfaction/contentment) if the anticipated reward is never unexpectedly absent (Burman et al., 2011; Moe et al., 2009; Spruijt et al., 2001).

This study investigated differences in behavioural measures that could suggest underlying emotional states between horses trained using PR and NR. We also investigated if these two training methods could induce more long-term changes in emotional states (moods). We hypothesized that training with PR would trigger positive emotions linked to the anticipation and consumption of food rewards (e.g. Moe et al., 2009). By contrast, we hypothesized that NR would trigger negative emotions, or at least less positive emotions than PR, as this training method involves the cessation of a negative stimulus, which is presumably less positively rewarding than a food reward. Then, we tested horse mood using a spatial judgment task (Burman et al., 2008a, 2009). Given that mood states arise as a result of an accumulation of shorter-term emotions (Mendl et al., 2010a), we also expected that horses trained with PR would show more optimistic bias, indicating more positive mood than horses trained with NR. Knowledge about the effects of training methods on the mental health of domesticated animals could add a new dimension to animal welfare, in order to promote not only better housing systems, but also better ways to work with animals.

2. Methods

2.1. Subjects and management conditions

The study was carried out at the Swiss National Stud Farm, Avenches, in February 2012. Subjects were 12 mares of three different breeds (three Franches-Montagnes, three Trotters and six Swiss half-bred) and aged 9–20 years (Table 1). They had been at the National Stud Farm for at least two years and they were not used for riding. They were all previously owned horses that retired to breeding, after withdrawal from sport. They were used for breeding as surrogate (foster) mares, but were not pregnant or manipulated for breeding purpose during the time of the study.

At all times, the 12 mares, which were split into two treatment groups (PR and NR, see Section 2.4), were housed all together in a large group housing system, which could normally contain up to 14 horses, with several compartments. Routine care of the study animals was provided by employees of the National Stud Farm, who were blind to the purpose of the study and to the treatment groups. They were fed in foraging stalls with haylage and a mix of corn and oats twice per day, 1 h before the experiments in the morning, and after the experiments in the evening. They had ad libitum access to straw in a rack and in laying areas, as well as water through an automatic water dispenser. Mares had access to an outside field 2–3 times per week depending on weather conditions. All subjects were lead from their home pen to the paddocks used for the treatment and judgment bias on a halter and rope.

Table 1. Characteristics of the horses used in the experiment (age and breed; FM = Franche-Montagne, TR = Trotter; SHB = Swiss half-bred), along with the treatment group (PR = positive reinforcement, NR = negative reinforcement). The side (L = left, R = right) on which they were rewarded during the judgment bias experiment (positive location) is also indicated (rewarded side).

Mare	Treatment group	Rewarded side	Age	Breed
1	PR	R	11	SHB
2	PR	R	12	SHB
3	PR	R	17	SHB
4	PR	L	9	TR
5	PR	L	12	FM
6	PR	L	20	SHB
7	NR	R	11	FM
8	NR	R	12	TR
9	NR	R	13	SHB
10	NR	L	10	FM
11	NR	L	15	SHB
12	NR	L	15	TR

Table 2. Description of the five phases of the study, along with duration and aims.

	Phases	Duration (day)	Aim
1	Habituation to the judgment bias set-up	1	Habituation
2	Training phase of the judgment bias experiment	3	Discrimination between rewarded and non-rewarded locations
3	Treatment	5	Inducing different moods using negative or positive reinforcement
4	Reminder session for the judgment bias experiment	1	Reminding reference locations
5	Testing phase of the judgment bias experiment	2	Assessing horse mood following treatment

2.2. Experimental design

The experiment consisted of several phases (Table 2): (1) the habituation phase (described in Section 2.3.2), aimed at habituating the horses to the judgment bias setup; (2) the judgment bias training phase (described in Section 2.3.3), during which the mares learned to discriminate the side where they were rewarded with food ("positive location"; left or right depending on the horse) from the non-rewarded side ("negative locations"; opposite direction); (3) the treatment (described in Section 2.4), during which mares were trained to perform several exercises using either positive or negative reinforcement; (4) a reminder session for the judgment bias test (described in Section 2.3.4); and (5) the judgment bias testing phase (described in Section 2.3.4), during which mares were exposed to ambiguous locations situated between the positive and negative reference locations.

2.3. Judgment bias

We used a judgment bias test, which uses spatial location as a stimulus (Burman et al., 2008a, 2009), to assess horse mood following treatment.

Fig. 1. Experimental setup for the judgment bias experiment. Position of the positive bucket (rewarded; right or left depending on the horses), the negative bucket (non-rewarded; opposite side as the rewarded one), the ambiguous buckets (equidistant angles between the positive and negative buckets), the waiting, start, central and end areas. The latency to reach the location was measured as the time to go from the starting line to the stop line next to the tested bucket.

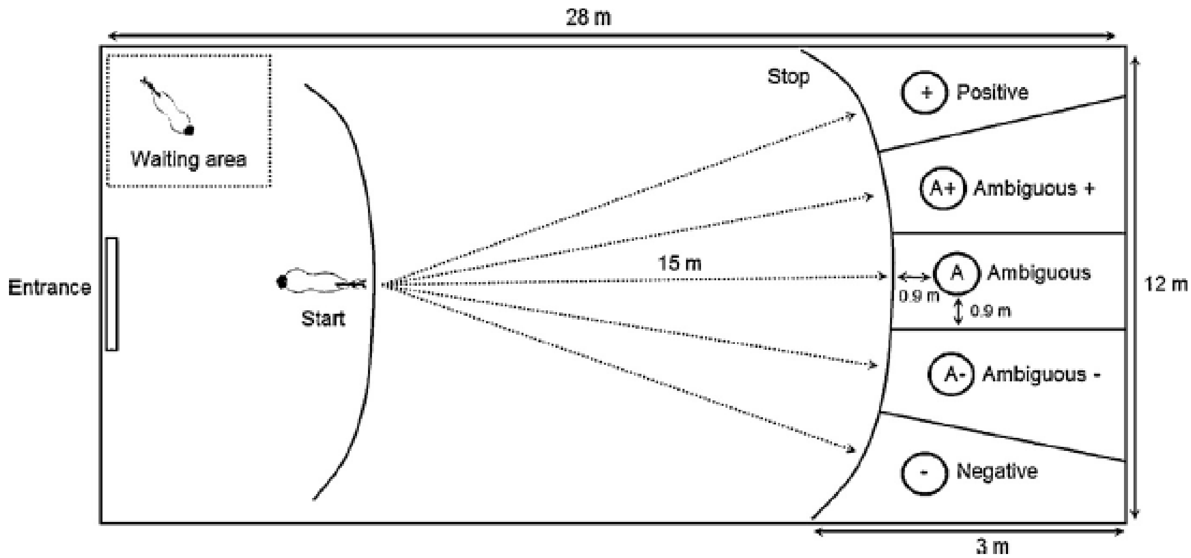
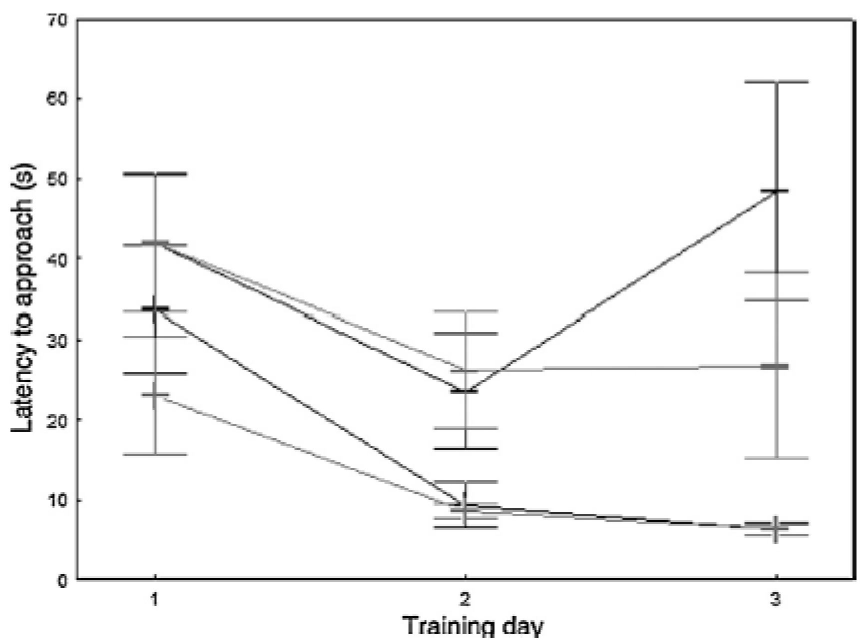


Fig. 2. Results of the judgment bias training phase. Latency to reach the rewarded location (+) and the non-rewarded location (–) over the three days of training for the treatment group subsequently trained with positive reinforcement (black) and the group subsequently trained with negative reinforcement (grey; mean ± SE per group and training day; n = 6 horse per group).



2.3.1. Experimental setup

The judgment bias experimental setup was adapted from Burman et al. (2008a) and Doyle et al. (2011, 2010). Experiments were carried out in a paddock with rubber mats on the ground (28 m × 12 m; Fig. 1). The horses had to cross the paddock, in order to reach a black bucket (50 cm wide and 30 cm deep), which always contained food (carrots, corn and oat) when it was situated at the positive location (always rewarded), or never contained food when it was situated at the negative location or at the ambiguous locations (never rewarded). The ambiguous locations were situated at intermediate distances between the positive and negative locations (ambiguous +, middle and ambiguous −; Fig. 1). They were never rewarded in order to avoid any association between the location and the presence of a food reward (Burman et al., 2009). The paddock was split in three main areas, which had been traced on the ground using a coloured spray: (1) an 8 m long start area delimited by the starting line; (2) a 15 m long central area that the horses had to cross to reach the bucket; and (3) a 3 m long end area, delimited by a stop line and divided into five smaller areas corresponding to the various locations (positive, ambiguous +, middle, ambiguous −, negative). The bucket was placed 90 cm behind the stop line and 90 cm from the right and left lines delimitating each of the five areas corresponding to the test locations (Fig. 1). At any of the locations, the bucket was thus situated at 15.9 m from the middle of the starting line, where the horse was placed at the beginning of the trial. The bucket was always covered with a 75 cm long and 1.5 cm thick wooden lid, in order to prevent any visual and olfactory cues indicating the location of the food reward (Briefer and McElligott, 2013).

2.3.2. Habituation

On the first day of the experiments, horses were habituated individually to the experimental paddock twice, once in the morning and once in the afternoon (two habituation sessions per horse). They were taken individually from their home enclosure to the paddock on a halter and rope. They were walked around the paddock and released there for 2 min. The black bucket used for the judgment bias experiment was placed in the start area (Fig. 1) and the horse was allowed to eat from it. The experimenter then placed the wooden lid on the bucket and allowed the horse to eat again, which enabled them to learn how to remove the lid in order to access the food. During the following 20 min (morning session) or 10 min (afternoon session), the horse was walked by the experimenter on a halter and rope to the other end of the paddock and released to allow the horse to go back to the bucket, remove the lid and feed.

2.3.3. Judgment bias training

The judgment bias training lasted three days and consisted of training the mares to discriminate between the positive and negative locations situated on each side of the paddock. Half of the horses assigned to the PR treatment group (n = 3 horses) and half of the horses assigned to the NR treatment group (n = 3 horses) were trained to expect food on the left (positive side = left of the end area). The other half (n = 3 PR and 3 NR mares) were trained to expect food on the right (positive side = right of the end area; Table 1; Fig. 1). The training procedure was adapted from Burman et al. (2009) and Briefer and McElligott (2013). On each training day, the mares received one session of six training trials: three positive and three negative. To facilitate learning, during the first session of six training trials (first training day), all mares received two positive trials, followed by two negative trials, then one positive and one negative (i.e. + + – – +–). During the following sessions, we used a pseudo-random sequence with no more than two consecutive positive or negative trials, and with the same number of positive and negative trials per session (e.g. + + – + – – , – + + – – + or + – – + + –). Each mare from the PR group was randomly paired with a mare from the NR group and these two mares received the same sequence over the three training days. The training ended after three days, when a significantly shorter latency to reach the positive location than the negative one had been obtained during two following days (Fig. 2).

2.3.4. Judgment bias testing

The judgment bias testing took place after the treatment phase (PR and NR; described in Section 2.4). The day after the end of treatment phase, and prior to the judgment bias test, each horse received one session of four trials, as a reminder. During this session, all mares received two positive trials, followed by two negative trials (i.e. + + – –), to facilitate memory. The judgment bias testing phase was carried out on the two following days. On each testing day, each mare received one session of seven testing trials. Each session consisted of two trials with the positive location, two trials with the negative location, and one trial with each of the ambiguous location. Following Briefer and McElligott (2013), each session started with one trial with the positive location and one trial with the negative location, or the opposite, as a reminder. Then, mares were tested with the three ambiguous locations in a random order, interspersed by the positive and the negative locations. For each mare, the order of testing with the ambiguous, positive and negative locations was counterbalanced over the two days, so that each ambiguous location ("A") would be tested one day after the positive location, and the other day after the negative location (i.e. day 1 = + − A3 + A2 − A1; day 2 = − + A1 − A2 + A3). Pairs of mares from the PR and NR groups, which had been allocated to the same sequence of trials during the training phase, also received the same sequence over the two testing days.

2.3.5. Judgment bias training and testing procedure

Two experimenters were present during the training and testing phases. The first experimenter (Exp 1), who was blind to the side at which subjects were rewarded (right or left), walked the horse on a halter and rope from the home pen to the paddock, while the second experimenter (Exp 2) prepared the first trial by placing the bucket with or without food, covered with a lid, at the appropriate location for the trial. Two black buckets, identical in size and shape, were used for the experiment; one bucket was filled with food and used for all the positive trials and one bucket was empty and used for all the negative and ambiguous trials. For each horse, the positive bucket was fully filled with food at the beginning of the session and the food quantity lasted for the whole session. For each trial, Exp 2 always presented only one of the two buckets at the appropriate location, while the other bucket was placed outside the paddock, out of sight of the tested horse. Exp 1 then entered the paddock, released the horse in the middle of the start line and stood still during the whole time, head down, while the horse was going away from her, towards the buckets (Fig. 1). Both experimenters then waited for the horse to access and eat a few bites of the food (positive location) or to reach the appropriate area where the bucket was located (negative and ambiguous locations). A 2 min intertrial interval followed, during which Exp 1 returned the horse to the waiting area (Fig. 1). While the horse was facing the opposite side, Exp 2 placed the bucket with or without food at the appropriate location for the next trial. During each trial, Exp 2 scored the latency to reach the bucket as the time from when one of the horse's front legs passed the start line, until one of its front legs passed one of the lines (stop line or lines on each side of the bucket) delineating the appropriate area where the bucket was located (Fig. 1). If the horses did not reach the area where the bucket was situated, it was brought back to the start area after 3 min by Exp 1, and the training/testing session continued. This happened on 6/108 occasions during a positive trial on the first (five occasions) or second (one occasion) training day, 13/108 occasions during negative training trials, 17/48 occasions during negative testing trial and 3/12 occasions when testing the ambiguous location next to the negative one. During the first training day, if the horse did not reach the positive location, Exp 1 brought the horse to the bucket and showed it the food. At the end of the daily session of 6 (training) or 7 (testing) trials, the subject was taken back to the home enclosure by Exp 1, while Exp 2 was refilling the positive bucket for the next horse.

2.4. Treatments

2.4.1. Training exercises

To study the effect of PR or NR on horse mood, the horses were trained, during five days to perform several exercises using only PR for one group (“PR group”) and NR for the other group (“NR group”). The training was carried out in a different paddock than the one used for the judgment bias experiment. It was situated close to the home enclosure, in order to avoid stress linked to isolation. The treatment started on the day after the judgment bias training phase and lasted for five days, after which the reminder and the judgment bias testing phase were carried out (Table 2). Each day, horses received three sessions (days 1 and 2) or two sessions (days 3–5) of 5 min each, interspersed by 2.5 min breaks (12.5–20 min per horse each day including breaks). Each day, all the horses from one group were trained, followed by all the horses from the other group, in a counterbalanced order (PR–NR or NR–PR depending on the days). All the horses were trained by the same experimenter (Exp 2), while Exp 1 was in charge of moving the horses to the training paddock and back, and collecting the data during the training.

The exercises used for the treatment were designed to be easily trained using either PR or NR. Exp 2 trained all horses to perform the same following sequence of exercises on demand, while standing afoot on the left side of the horse;

1. Stand still next to the trainer, with the head forward.
2. Walk next to the trainer.
3. Stop next to the trainer.
4. Move backward.
5. Flex the neck alternatively on the right and on the left.
6. Lower the head down towards the ground.
7. Walk on a tarp (1.5 m × 2.5 m) that was covered by plastic bottles.
8. Climb on a wooden podium (70 cm large × 70 cm long × 50 cm high).
9. Push a blue balloon (diameter 1 m) with the nose or leg.
10. Lift one of the front legs upward.

Each day, this sequence of exercises was trained or rehearsed (when successfully learned) in the same order for all horses. Each exercise was trained during 5 min maximum, unless well executed, before moving to the next one, and so on, until the end of the daily training session. An exercise was considered to be well executed when easily performed on demand. Depending on individual performances, and because the daily training sessions were limited in duration for each horse, the number of exercises performed on each of the training day could differ between horses. A daily session always ended up with the execution of the last exercise in the sequence that was correctly performed by the horse.

2.4.2. Positive and negative reinforcement

All horses were trained individually, unmounted, on a halter and rope. The same procedure was applied for both PR and NR groups (Heleski et al., 2008; Hendriksen et al., 2011; Innes and McBride, 2008; Sankey et al., 2010a); each movement towards a correct execution of the exercise was immediately rewarded (“shaping procedure”; McGreevy and McLean, 2010). For the PR group, the reward was a click using a clicker, followed by a piece of food (food pellet for horse; “clicker method”; Kurland, 1999). To guide their movement, PR horses were trained to follow a target (55 cm long stick, with a tennis ball at the end), by initially rewarding them when they were touching the target with their nose. Conversely, for the

NR group, the reward was the cessation of an uncomfortable stimulus. The stimulus was a gentle pressure on the halter, gently shaking or pulling the lead or/and a gentle pressure on the horse's flank using a stick, depending on the exercise. The intensity of the stimulus was gradually increased if the horse was not responding to it or not responding correctly. As soon as the expected response was shown, the experimenter stopped the stimulus immediately and the horse could rest for a short break (5 s) before continuing the training. The movements of NR horses were guided using such gradual pressure intensity, which was removed as soon as they moved in the correct direction. Exp 2 followed a detailed training procedure that was identical for all horses of a given group. The two groups received the same amount of training (mean ± SE total training time per horse and per exercise (exercises 1–9): PR, 485.56 ± 52.73 s; NR, 441.02 ± 37.01 s; linear mixed-effect model, with horse identity nested within exercise number as a random factor, and group as a fixed factor: $F_{1,84} = 0.81$, $p = 0.37$).

2.4.3. Behavioural measures

All training sessions were video recorded using a camcorder (Sanyo VPC-WH1, Waterproof, High-Definition). From the videos, the following behaviours were scored, based on a 1–3 scale (Table 3), and using an instantaneous time sampling every 10 s; body tension, attitude towards the trainer, head level and ear position. Exercises varied in duration between horses, and also within horse between the training days. Thus, we calculated the percentage of time each horse was given a score of 1, 2 or 3 on the established scale for each behaviour, by dividing the occurrence of each score by the number of 10 s samples for the whole exercise (e.g. two occurrences of score 1, one occurrence of score 2 and 44 occurrences of score 3 = 4.3% for score 1, 2.1% for score 2 and 93.6% for score 3). For each parameter, the scale was designed as a continuum between a behaviour suggesting a negative emotional state (score of 1 for body tension, attitude towards trainer and ear position, and score of 3 for head level) and a positive emotional state (score of 3 for body tension, attitude towards trainer and ear position, and of score 1 for head level). Scores of 2 for all behaviours indicated an intermediate between negative and positive states (Table 3). Therefore, we calculated a general score per horse for each of these behaviours as follows: (% score 1 × 1) + (% score 2 × 2) + (% score 3 × 3). According to this equation, a horse would obtain a general score of 300 if it always scored 3 for one behaviour, a general score of 200 if it always scores 2, and a general score of 100 if it always scores 1 (range = 100–300; for the example above, the general score would be (4.3 × 1) + (2.1 × 2) + (93.6 × 3) = 289.4). Higher scores for body tension (more motivation), for attitude towards trainer (more contact with the trainer) and for ear position (ears more forward) thus suggested a more often positive emotional state, while higher scores for head level (head higher) suggested a more often negative emotional state (Briefer Freymond et al., 2013; McDonnell and Haviland, 1995; Rietmann et al., 2004; Visser et al., 2009).

2.5. Data analyses

For the behavioural measures taken during the treatment phase (training exercises), the analyses were carried out on the average values for each subject and for each exercise. All horses reached exercise 7, eight horses from both groups (three PR and five NR) reached exercise 8, and two NR mares reached exercise 9 (exercise reached per horse: range = exercise 7–9; mean = exercise 7.83 ± 0.21; n = 12 horses). We therefore decided to focus on behavioural measures collected during exercises 1–8, which contained data from both groups, for our analyses.

For the judgment bias training phase, the analyses were carried out on the average latency for each subject to reach the positive and negative location on each of the three training day. The data on the reminder day (day after the treatment and before starting the test; Table 2) were analysed separately. For the judgment bias testing phase, the analyses were carried out on the latency to reach each of the locations (ambiguous, positive and negative), averaged over the two test days for each subject. If a mare did not reach the area where the bucket was situated during the training or testing phases of the judgment bias experiment, we attributed, for the trial, a latency corresponding to the maximum time taken by this individual, over the days of training and testing, to reach any of the locations. This allowed us to avoid replacing these values by either missing data or by an artificial maximum of 3 min that had been decided by us during the planning of the experiments. It also maintained the individual variability in latencies to reach the locations (Briefer and McElligott, 2013).

We analysed behavioural measures from the treatment (training exercises), and the latency data from the training and testing phases of the judgment bias using linear mixed-effects models (LMM; lme function in R; Bates, 2005). This allowed us to investigate or control for the effect of several factors (age, breed, training side, training day, group, location). For the treatment phase, the initial models for the behavioural measures included body tension score, attitude towards trainer score, head level score or ears position score calculated for the horses as a response variable (four different models), as well as the breed (Franches-Montagnes, Trotters and Swiss half-bred) and age (9–20) of the mares as fixed factors to control for age and breed differences. The training group (PR or NR) and the exercise number (Ex1–Ex8) were included with the interaction between them, as fixed factors, to test for their effect. For the judgment bias experiment, models for both the training and testing phases included the latency to reach the various locations as a response variable, as well as the age, breed and rewarded side (left or right) of the mares as fixed factors, in order to control for age and breed differences and for potential laterality bias (e.g. De Boyer Des Roches et al., 2008). The day of training (1–3; for the training phase only), the location (positive and negative for the training phase; positive, negative, ambiguous +, middle and ambiguous − for the testing phase) and the group were included, with all possible interactions between them, as fixed factors, in order to test for their effect. Finally, for all models, the identity of the mares was included as a random factor to control for repeated measurements of the same subjects.

We then removed non-significant terms using a standard model simplification procedure. Each nonsignificant term was removed if the deletion did not cause any significant reduction in goodness of fit. The two models with and without each term, both fitted with the maximum likelihood method (ML), were compared using a likelihood ratio test. We present the results after model simplification and with restricted maximum likelihood method (RELM). When the interaction effect between two factors was significant, further post hoc comparisons were carried out using LMM including control factors that remained in the final models. The residuals of the models were checked graphically for normal distributions and homoscedasticity, and were log-transformed if necessary (body relaxation score, attitude towards trainer score and latencies during the reminder day and during testing). Statistical analyses were carried out using R v. 2.15.0 (R Development Core Team, 2012). The significance level was set at $\alpha = 0.05$. No Bonferroni correction was applied for the post hoc comparisons due to the small sample sizes (Nakagawa, 2004). All means are given with standard errors.

2.6. Ethics

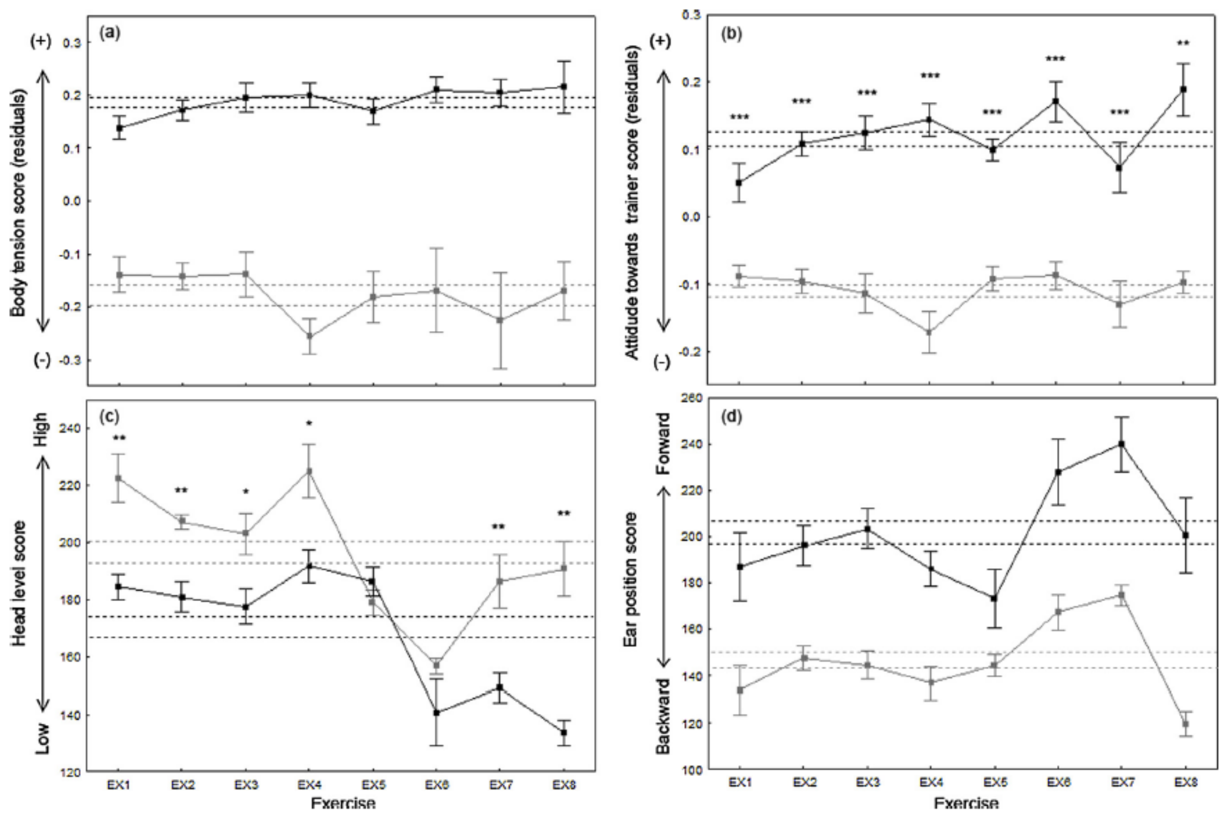
Animal care and all experimental procedures were in accordance with the International Society for Applied Ethology guidelines.

3. Results

3.1. Judgment bias training

Mares went faster to the positive (rewarded) location than to the negative (non-rewarded) location, and thus successfully learned the task (Fig. 2). There was no significance difference between the two treatment groups (PR and NR) in their latencies to reach the locations (Fig. 2).

Fig. 3. Behavioural responses to the treatment. (a) Body tension score (residuals of log-transformed scores controlled for age and breed), (b) attitude towards trainer score (residuals of log-transformed scores controlled for age and breed), (c) head level score and (d) ear position score measured during the training exercises 1–8 with positive reinforcement (black) and negative reinforcement (grey; mean ± SE per group and per exercise; 12 horses for exercises 1–7 and 8 horses for exercise 8). More positive body tension score's residuals indicated a more positive tension (+) and more negative residuals a more negative/indifferent tension (–). Similarly, for the attitude towards trainer score, more positive residuals indicated a more positive attitude (+) and more negative residuals a more negative attitude (–). High head level scores indicated a higher position of the head and lower scores a low position. Finally, high ear position scores indicated that the ears were more often forward and low scores that the ears were more often backward. Differences between the two treatment groups (negative and positive reinforcement) for each exercise are indicated as follows (when the interaction term between group and exercise was significant): *p < 0.05; **p < 0.01; ***p < 0.001; non-significant otherwise (linear mixed-effect models). Dotted lines indicate mean ± SE across the exercise for horses trained with negative reinforcement (grey) and positive reinforcement (black).



The model selection procedure for the training session revealed an effect of day of training on the latencies to reach the locations (LMM: $F_{1,58} = 6.30$, p = 0.015; Fig. 1). The general latency to reach the locations changed over the training phase (latency: day 1 = 35.16 ± 4.10 s; day 2 = 16.95 ± 2.98 s; day 3 = 21.94 ± 5.51 s; n = 12 horses). The effect of the treatment group was not significant and was removed during model selection (likelihood-ratio test: $\chi^2_1 = 0.37$, $p = 0.54$). The latencies of PR mares were therefore not different to the latencies of NR mares during training (Fig. 2). There was a general effect of location on the latencies (LMM: $F_{1,58} = 21.97$, $p < 0.0001$); mares reached the positive location substantially faster (latency = 14.56 ± 2.46 s) than the negative location (latency = 34.81 ± 4.00 s; $n = 12$ horses; Fig. 2).

The control factors (rewarded side, age and breed) and the interaction terms included in the initial model that are not mentioned above (day–location–group; location–group; day–group; day–location) did not have significant effects on the latencies to reach the locations. They were thus removed during model selection.

On the reminder day, horses could still correctly discriminate between the two reference locations, as shown by the significantly shorter latency to reach the positive location (5.57 ± 0.41 s) than the negative one (35.38 ± 10.97 s; $n = 12$ horses; LMM: $F_{1,11} = 17.04$, $p = 0.0017$). The effect of the treatment group was not significant and was removed during model selection (likelihood-ratio test: $\chi^2_1 = 1.74$, $p = 0.19$). The control factors (training side, age and breed) and the interaction term between group and location were not significant either and were also removed during model selection.

3.2. Behavioural measures during treatment

Differences between treatment groups were consistent for all variables; PR mares showed a more motivated body tension, a more positive attitude towards the trainer, a lower head position and more forward ears than NR mares (Fig. 3).

3.2.1. Body tension

There were no significant differences between the two treatment groups, depending on the type of exercise, for the body tension score (LMM: interaction effect, $F_{7,66} = 1.88$, $p = 0.086$). However, this interaction term could not be removed during the model selection, because doing so caused a significant reduction in goodness of fit (likelihood-ratio test: $\chi^2_1 = 14.54$, $p = 0.042$). The model selection procedure did not reveal any significant variation in body tension between the exercises (LMM: $F_{7,66} = 0.73$, $p = 0.64$; Fig. 3a). There was a general effect of the group, with PR mares having higher body tension scores, indicating more motivated body tension, than NR mares (LMM: $F_{1,7} = 246.64$, $p < 0.0001$; Fig. 3a). The two control factors had a significant effect on body tension scores (LMM: age, $F_{1,7} = 6.38$, $p = 0.040$; breed, $F_{2,7} = 11.58$, $p = 0.006$).

3.2.2. Attitude towards trainer

There was a significant interaction effect between exercise and treatment group for the attitude towards trainer score (LMM: $F_{7,66} = 4.70$, $p = 0.0003$). Post hoc comparisons revealed that PR mares had higher attitude towards trainer scores, indicating a more positive attitude than NR mares, for all exercises (Fig. 3b). Concerning the control factors, breed had a significant effect on attitude towards trainer scores (LMM: $F_{2,7} = 30.10$, $p = 0.0004$). Age did not have a significant effect in the final model (LMM: $F_{1,7} = 2.40$, $p = 0.16$), but this term could not be removed during the model selection, because its deletion caused a significant reduction in goodness of fit (likelihood-ratio test: $\chi^2_1 = 4.39$, $p = 0.036$).

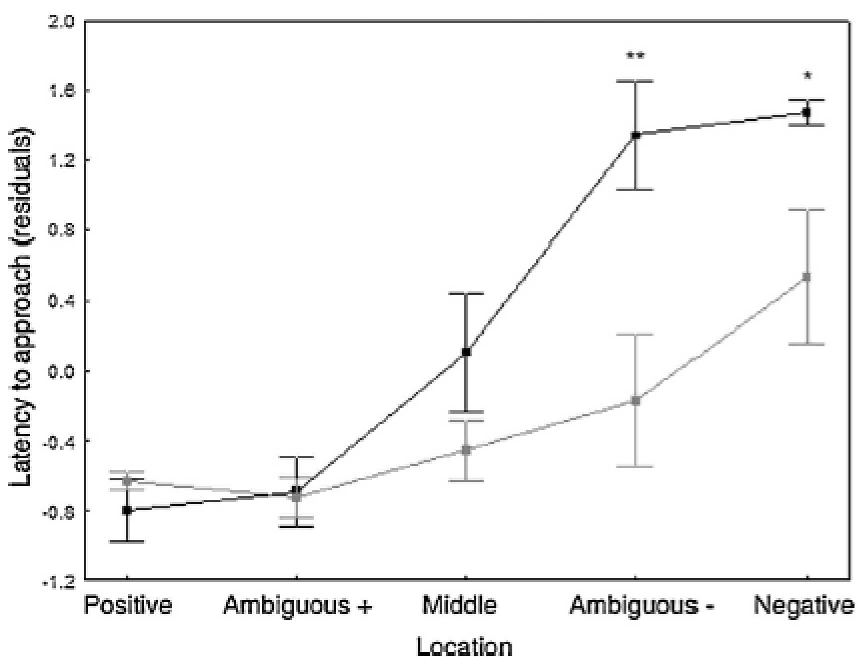
3.2.3. Head level

There were significant differences between the two groups depending on the type of exercise for head level scores (LMM: interaction term, $F_{7,66} = 5.48$, $p = 0.0001$). Post hoc comparisons revealed that the head level scores of PR mares were significantly lower than the scores of NR mares for all exercises, except exercises 5 (flex the neck) and 6 (lower the head; Fig. 3b). Age and breed (control factors) did not significantly affect head level scores and they were removed during model selection.

3.2.4. Ear position

The interaction term between treatment group and exercise was not significant and was removed during model selection for ear position score (likelihood-ratio test: χ^2_1 = 9.16, p = 0.24). Ear position scores varied significantly between exercises (LMM: $F_{7,73}$ = 8.68, p < 0.0001; Fig. 3d). There were significant differences between the two groups; The ears of PR mares were forward more often than the ears of NR mares (LMM: $F_{1,10}$ = 88.69, p < 0.0001; Fig. 3d). The two control factors, age and breed did not significantly affect ear position scores and were removed during model selection.

Fig. 4. Results of the judgment bias experiment. Latency (residuals of log-transformed latency controlled for age) to reach the five locations (see Fig. 1 for details) during the two days of test for the treatment group formerly trained with positive reinforcement (black) and the group trained negative reinforcement (grey; mean ± SE per group per location; *n* = 6 horses per group). More positive residuals indicated longer latencies. Differences between the two groups (negative and positive reinforcement) for each location are indicated as follows: **p* < 0.05; ***p* < 0.01; non-significant otherwise (linear mixed-effect models).



3.3. Judgment bias testing

The results of the testing phase of the judgment bias experiment revealed that mares reached the ambiguous locations between the positive and negative locations with intermediate latencies (Fig. 4). PR mares were generally slower than NR mares to reach locations; they went significantly slower than NR mares to the negative location and the adjacent ambiguous location (Fig. 4).

The interaction effect between group and location was significant (LMM: $F_{4,40}$ = 4.19, p = 0.006), indicating differences between the groups depending on the locations. Post hoc comparisons revealed that PR mares reached the negative location and the ambiguous location next to it (“ambiguous –”) slower than NR mares. There was no significant difference between the latencies taken by PR and NR mares to reach the other locations (Fig. 4).

Age did not significantly affect the latencies to reach the locations (LMM: $F_{1,9} = 2.14$, $p = 0.18$), but this term could not be removed during model selection, because doing so caused a significant reduction in goodness of fit (likelihood-ratio test: $\chi^2_1 = 3.86$, $p = 0.049$). The other control factors (breed and training side) included in the initial model did not have a significant effect on the latencies to reach the locations, and were removed during model selection.

4. Discussion

Knowledge about the effect of training methods on the emotional states of domesticated animals could help improve welfare, not only in terms of how animals are housed, but also how people interact with animals. We assessed behavioural measures that could suggest underlying emotional states during training sessions using either positive or negative reinforcement in horses. We also tested if the training method (PR or NR) had a long-term impact on horse emotional states, using a judgment bias experiment. As predicted, behavioural measures suggested that mares trained with PR experienced more positive emotions (i.e. more motivated, more positive attitude towards the trainer, head lower and ears more forward), than mares trained with NR (i.e. more tensed, more negative/indifferent attitude towards the trainer, head higher and ears more backwards; McDonnell and Poulin, 2002; McDonnell and Haviland, 1995; Rietmann et al., 2004; Sankey et al., 2010a, 2010b; Visser et al., 2009). However, during the judgment bias test, NR mares were then more optimistic compared to PR mares, despite their behaviour suggesting more negative emotions during training. NR mares could have been more motivated to obtain food rewards than PR mares, which had been rewarded throughout the training exercises (Burman et al., 2011). Alternatively, NR mares could have developed an optimistic bias, suggesting more positive moods, triggered by a release from stress experienced during the training exercises (Briefer and McElligott, 2013; Doyle et al., 2010; Sanger et al., 2011). This first attempt to measure judgment biases in horses suggests that this is a promising method to assess horse optimism.

4.1. Behavioural indicators of emotions during treatment

Behavioural indicators of emotions were remarkably different between the two treatment groups. Body tension scores were lower, indicating more motivation in PR mares, whereas it indicated more stress and tension/lack of motivation in NR mares, for all exercises. Our measures of attitude towards the trainer showed that, for all exercises, PR mares tended to establish contact with the trainer, whereas NR mares tended to avoid contact with/be indifferent towards the trainer. PR mares had a lower head level than NR mares for all exercises, except exercises 5 and 6, which consisted of flexing the neck and lowering the head and resulted obviously in similar head levels for both groups. Finally, the ears of PR mares were more forward than the ears of NR mares, for all exercises. All these measures suggest that PR mares were experiencing more positive emotions than NR mares during the treatment (Boissy et al., 2011; Reefmann et al., 2009; Sankey et al., 2010a, 2010b; Young et al., 2012), although some of our measures, such as the attitude towards the trainer, could indicate motivation for food more than underlying emotions. Similar differences have been found in horses trained with PR or NR in other studies. For instance, Sankey et al. (2010a) observed more forward ears, less head lifting, lower neck position and faster contact initiation with human in PR compared to NR horses trained to walk backward. Hendriksen et al. (2011) found less discomfort behaviour (eye and nostril widening, tail whipping) and less avoidance towards the training in PR than NR horses trained to enter a trailer. Innes and McBride (2008) showed that PR horses made more active contact with the trainer than NR horses. Finally, Warren-Smith and McGreevy (2007) found that horses trained with both PR and NR to halt, shook their heads vertically less than horses trained only with NR. All these results strongly suggest that PR could be associated with more positive emotions, more positive contact with the trainer, and thus better horse welfare at work than NR. PR could even be a more efficient form of training (i.e. higher success rate, faster) than NR (Hendriksen et al., 2011; Sankey et al., 2010a).

4.2. Judgment bias test

We found significant differences between the two groups in their responses to the judgment bias experiment. As mares from the two groups were all housed together in a large housing system, and as National Stud employees who were providing routine care of the animals were blind to the purpose of the study and to the treatment groups, it is highly likely that cognitive bias differences observed between the two groups resulted from our treatment (PR versus NR), and not from other confounding factors (e.g. housing, routing care, feeding). Despite displaying behavioural responses suggesting more negative emotions during training, NR mares went significantly faster than PR mares to the negative location and the ambiguous location next to it (ambiguous −). These results suggest that NR mares expected less negative events to occur and were therefore less pessimistic than PR mares. According to Mendl et al. (2010a), NR mares could have been in a mood appropriate to an intermediate reward opportunity environment, and were optimistic about the expectation of negative events (absence of food in our case), compared to PR mares.

Our results cannot be explained by PR horses being fully satiated, because the two groups were fed in the same way in their home enclosure, and were tested on the second day following the end of the treatment. Furthermore, we did not find any difference between groups in their latency to reach the positive location, suggesting no obvious difference in motivation to participate in the test (Burman et al., 2011). However, we found a difference between PR and NR horses in their latency to reach the negative reference location, despite this difference not being present during the reminder day. It is possible that NR mares differed in their feeding motivation compared to PR mares (Burman et al., 2011). NR mares, which were rewarded during the judgment bias test, but not during the treatment phase (NP and NR exercises), could have been more motivated to obtain food than PR mare. By contrast, PR mares, which obtained food only at the reference location during the judgment bias test, instead of throughout the treatment phase (reinforcer = clicker + food), could have been less motivated to check in the ambiguous and negative locations ("incentive contrast"; Flaherty, 1996; Burman et al., 2008b). This explanation does not necessarily imply underlying treatment group difference in mood. Similarly, non-rewarded dogs were shown to be more optimistic than rewarded dogs (Burman et al., 2011).

Alternatively, our treatments could have really induced different moods in PR and NR horses, but in the opposite direction to that expected. We predicted that, as a result of accumulation of short-term positive emotional states during treatment, PR horses would be experiencing positive moods. Similarly, we predicted that NR horses would be in a negative mood, following an accumulation of short-term negative emotional states. Yet, our results show more optimism in NR than PR mares, following negative emotions during treatment. If this optimism really indicates underlying moods, then NR mares were in a more positive mood than PR mares. Similar unexpected optimism has been found in sheep following stressful shearing and restraint (Doyle et al., 2010; Sanger et al., 2011), and in female goats rescued from poor welfare (Briefer and McElligott, 2013). This optimism has been suggested to be triggered by a release from a stressful situation (negative treatment; Doyle et al., 2010), and could be present in females more than males (Briefer and McElligott, 2013). It could also be compared to optimistic judgment bias following a change in environment, as shown in pigs (*Sus scrofa*) and rats, which were transferred from an unenriched to an enriched environment (Brydges et al., 2011; Douglas et al., 2012). Animals that have been stressed might seek positive events to balance their situation once a negative event ceases (Spruijt et al., 2001). Therefore, our results suggest that release from the negative emotions experienced during NR training induces optimism, at least in female horses. We propose that this effect could be quite common; when a short-term (Doyle et al., 2010; Sanger et al., 2011), or long-term stressor ends (Briefer and McElligott, 2013), females might develop optimistic bias. After a particularly long-term stressor (i.e. lasting > 1 year), this optimistic bias could even last for years (Briefer and McElligott, 2013). Further judgment bias experiments testing reactions to ambiguous locations after the negative treatment ends could be compared to situations where the same negative treatment is still applied, in order to verify this hypothesis.

5. Conclusion

The judgment bias approach is a promising method for measuring optimism and pessimism in horses. The behaviour of horses during training using NR suggests more negative emotions than when using PR. Yet, horses trained using NR develop more optimistic bias than horses trained using PR, after the training stops. This optimistic bias could be due to higher feeding motivation in NR compared to PR mares, or to the cessation of a negative event. We therefore recommend training method using only PR, or combining NR and PR (Warren-Smith and McGreevy, 2007). Training methods that trigger good mental health should be as important for animal welfare as appropriate housing systems.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgements

We are grateful to H. Zurkinden, to the vet team of ISME at Agroscope Swiss National Stud Farm, and to all the staff of Agroscope Swiss National Stud Farm for assistance. We are also grateful to S. Hintze and A. McElligott for helpful comments on the manuscript. E. Briefer is funded by a Swiss National Science Foundation (PZ00P3 148200) fellowship.

References

Bates, D., 2005. Fitting linear mixed models in R. R News 5, 27–30.

Bateson, M., Matheson, S.M., 2007. Performance on a categorisation task suggests that removal of environmental enrichment induces pessimism in captive European starlings (Sturnus vulgaris). Anim. Welf. 16, 33–36.

Bateson, M., Desire, S., Gartside, S.E., Wright, G.A., 2011. Agitated honeybees exhibit pessimistic cognitive biases. Curr. Biol. 21, 1070–1073.

Boissy, A., Manteuffel, G., Jensen, M.B., Moe, R.O., Spruijt, B., Keeling, L.J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., Aubert, A., 2007. Assessment of positive emotions in animals to improve their welfare. Physiol. Behav. 92, 375–397.

Boissy, A., Aubert, A., Désiré, L., Greiveldinger, L., Delval, E., Veissier, I., 2011. Cognitive sciences to relate ear postures to emotions in sheep. Anim. Welf. 20, 47–56.

Briefer, E.F., 2012. Vocal expression of emotions in mammals: mechanisms of production and evidence. J. Zool. 288, 1–20.

Briefer, E.F., McElligott, A.G., 2013. Rescued goats at a sanctuary display positive mood after former neglect. Appl. Anim. Behav. Sci. 146, 45–55.

Briefer Freymond, S., Briefer, E.F., Niederhäusern, R.V., Bachmann, I., 2013. Pattern of social interactions after group integration: a possibility to keep stallions in group. PLoS ONE 8, e54688.

Brilot, B., Asher, L., Bateson, M., 2010. Stereotyping starlings are more pessimistic. Anim. Cogn. 13, 721–731.

Brydges, N.M., Leach, M., Nicol, K., Wright, R., Bateson, M., 2011. Environmental enrichment induces optimistic cognitive bias in rats. Anim. Behav. 81, 169–175.

Burman, O.H.P., Parker, R., Paul Elizabeth, S., Mendl, M., 2008a. A spatial judgement task to determine background emotional state in laboratory rats, Rattus norvegicus. Anim. Behav. 76, 801–809.

Burman, O.H.P., Parker, R.M.A., Paul, E.S., Mendl, M., 2008b. Sensitivity to reward loss as an indicator of animal emotion and welfare. Biol. Lett. 4, 330–333.
Burman, O.H.P., Parker, R.M.A., Paul, E.S., Mendl, M.T., 2009. Anxiety-induced cognitive bias in non-human animals. Physiol. Behav. 98, 345–350.
Burman, O., McGowan, R., Mendl, M., Norling, Y., Paul, E., Rehn, T., Keeling, L., 2011. Using judgement bias to measure positive affective state in dogs. Appl. Anim. Behav. Sci. 132, 160–168.
Dawkins, M.S., 2008. The science of animal suffering. Ethology 114, 937–945.
De Boyer Des Roches, A., Richard-Yris, M.-A., 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, 487–490.
Désiré, L., Boissy, A., Veissier, I., 2002. Emotions in farm animals: a new approach to animal welfare in applied ethology. Behav. Processes 60, 165–180.
Douglas, C., Bateson, M., Walsh, C., Bédué, A., Edwards, S.A., 2012. Environmental enrichment induces optimistic cognitive biases in pigs. Appl. Anim. Behav. Sci. 139, 65–73.
Doyle, R.E., Fisher, A.D., Hinch, G.N., Boissy, A., Lee, C., 2010. Release from restraint generates a positive judgement bias in sheep. Appl. Anim. Behav. Sci. 122, 28–34.
Doyle, R.E., Hinch, G.N., Fisher, A.D., Boissy, A., Henshall, J.M., Lee, C., 2011. Administration of serotonin inhibitor p-chlorophenylalanine induces pessimistic-like judgement bias in sheep. Psychoneuroendocrinology 36, 279–288.
Flaherty, C.F., 1996. Incentive Relativity. Cambridge University Press, Cambridge, UK.
Forkman, B., Boissy, A., Meunier-Salaün, M.-C., Canali, E., Jones, R.B., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiol. Behav. 92, 340–374.
Fureix, C., Jego, P., Henry, S., Lansade, L., Hausberger, M., 2012. Towards an ethological animal model of depression? A study on horses. PLoS ONE 7, e39280.
Hall, C., Goodwin, D., Heleski, C., Randle, H., Waran, N., 2008. Is there evidence of learned helplessness in horses? J. Appl. Anim. Welf. Sci. 11, 249–266.
Harding, E.J., Paul, E.S., Mendl, M., 2004. Animal behaviour: cognitive bias and affective state. Nature 427, 312.
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, e7625.
Heleski, C., Bauson, L., Bello, N., 2008. Evaluating the addition of positive reinforcement for learning a frightening task: a pilot study with horses. J. Appl. Anim. Welf. Sci. 11, 213–222.
Hendriksen, P., Elmgreen, K., Ladewig, J., 2011. Trailer-loading of horses: is there a difference between positive and negative reinforcement concerning effectiveness and stress-related signs? J. Vet. Behav. 6, 261–266.
Innes, L., McBride, S., 2008. Negative versus positive reinforcement: an evaluation of training strategies for rehabilitated horses. Appl. Anim. Behav. Sci. 112, 357–368.
Keltner, D., Lerner, J.S., 2010. Emotion. In: Handbook of Social Psychology. John Wiley & Sons, London, pp. 317–352.
Kurland, A., 1999. Clicker Training for Your Horse. Sunshine Books, Waltham, MA.
MacLeod, C., Byrne, A., 1996. Anxiety, depression, and the anticipation of future positive and negative experiences. J. Abnorm. Psychol. 105, 286–289.
Matheson, S.M., Asher, L., Bateson, M., 2008. Larger, enriched cages are associated with optimistic response biases in captive European starlings (Sturnus vulgaris). Appl. Anim. Behav. Sci. 109, 374–383.
McDonnell, S.M., Haviland, J.C.S., 1995. Agonistic ethogram of the equid bachelor band. Appl. Anim. Behav. Sci. 43, 147–188.
McDonnell, S., Poulin, A., 2002. Equid play ethogram. Appl. Anim. Behav. Sci. 78, 263–290.
McGreevy, P., McLean, A., 2010. Equitation Science. Wiley-Blackwell Publishing, United Kingdom.

Mendl, M., Paul, E.S., 2004. Consciousness, emotion and animal welfare: insights from cognitive science. Anim. Welf. 13, 17–25.

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. Appl. Anim. Behav. Sci. 118, 161–181.

Mendl, M., Burman, O.H.P., Paul, E.S., 2010a. An integrative and functional framework for the study of animal emotion and mood. Proc. R. Soc. Lond. B 277, 2895–2904.

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., Casey, R., 2010b. Dogs showing separation-related behaviour exhibit a pessimistic cognitive bias. Curr. Biol. 20, R839–R840.

Moe, R.O., Nordgreen, J., Janczak, A.M., Spruijt, B.M., Zanella, A.J., Bakken, M., 2009. Trace classical conditioning as an approach to the study of reward-related behaviour in laying hens: a methodological study. Appl. Anim. Behav. Sci. 121, 171–178.

Nakagawa, S., 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behav. Ecol. 15, 1044–1045.

Nettle, D., Bateson, M., 2012. The evolutionary origins of mood and its disorders. Curr. Biol. 22, R712–R721.

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

R Development Core Team, 2012. R Foundation for Statistical Computing, Vienna, Austria.

Reefmann, N., Wechsler, B., Gygax, L., 2009. Behavioural and physiological assessment of positive and negative emotion in sheep. Anim. Behav. 78, 651–659.

Rietmann, T.R., Stuart, A.E.A., Bernasconi, P., Stauffacher, M., Auer, J.A., Weishaupt, M.A., 2004. Assessment of mental stress in warmblood horses: heart rate variability in comparison to heart rate and selected behavioural parameters. Appl. Anim. Behav. Sci. 88, 121–136.

Sanger, M.E., Doyle, R.E., Hinch, G.N., Lee, C., 2011. Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. Appl. Anim. Behav. Sci. 131, 94–103.

Sankey, C., Richard-Yris, M.-A., Henry, S., Fureix, C., Nassur, F., Hausberger, M., 2010a. Reinforcement as a mediator of the perception of humans by horses (Equus caballus). Anim. Cogn. 13, 753–764.

Sankey, C., Henry, S., Górecka-Bruzda, A., Richard-Yris, M.A., Hausberger, M., 2010b. The way to man's heart is through his stomach: what about horses. PLoS ONE 5, e15446.

Scherer, K.R., 1984. On the nature and function of emotion: a component process approach. In: Approaches to Emotion. Erlbaum, Hillsdale, NJ, pp. 293–317.

Sharot, T., 2011. The optimism bias. Curr. Biol. 21, R941–R945.

Skinner, B.F., 1938. The Behavior of Organisms: An Experimental Analysis. Appleton-Century-Crofts, New York.

Spruijt, B.M., van den Bos, R., Pijlman, F.T.A., 2001. A concept of welfare based on reward evaluating mechanisms in the brain: anticipatory behaviour as an indicator for the state of reward systems. Appl. Anim. Behav. Sci. 72, 145–171.

Strunk, D.R., Lopez, H., DeRubeis, R.J., 2006. Depressive symptoms are associated with unrealistic negative predictions of future life events. Behav. Res. Ther. 44, 861–882.

Visser, E.K., van Dierendonck, M., Ellis, A.D., Rijksen, C., van Reenen, C.G., 2009. A comparison of sympathetic and conventional training methods on responses to initial horse training. Vet. J. 181, 48–52.

Waran, N., McGreevy, P.D., Casey, R.A., 2002. Training methods and horse welfare. In: The Welfare of Horses. Kluwer Academic Publishers, The Netherlands, pp. 151–180.

Warren-Smith, A.K., McGreevy, P.D., 2007. The use of blended positive and negative reinforcement in shaping the halt response of horses (Equus caballus). Anim. Welf. 16, 481–488.

Wright, W.F., Bower, G.H., 1992. Mood effects on subjective probability assessment. Organ. Behav. Hum. Decis. Process. 52, 276–291.
Young, T., Creighton, E., Smith, T., Hosie, C., 2012. A novel scale of behavioural indicators of stress for use with domestic horses. Appl. Anim. Behav. Sci. 140, 33–43.