Validation of a fear test in sport horses using infrared thermography

Dai Francesca[1*], Cogi Nathalie Hélène[1], Heinzl Eugenio Ugo Luigi[1], Dalla Costa Emanuela[1], Canali Elisabetta[1], Minero Michela[1]

[1]Dipartimento di Scienze Veterinarie e Sanità Pubblica, Università degli Studi di Milano, Via Celoria 10, 20133 Milano, Italy

e-mail: francesca.dai@unimi.it

nathalie.cogi@studenti.unimi.it

eugenio.heinzl@unimi.it

emanuela.dallacosta@unimi.it

elisabetta.canali@unimi.it

michela.minero@unimi.it

[*]Corresponding author:

Dai Francesca c/o Dipartimento di Scienze Veterinarie e Sanità Pubblica, Università degli Studi di Milano, Via Celoria 10, 20133 Milano, Italy.

e-mail francesca.dai@unimi.it

tel. +39 0250318033

Abstract

The aims of the present study were to assess feasibility and validity of a fear test in adult sport horses and to investigate whether the exposure to a fearful stimulus induces a change in eye temperature. Fifty horses, aged 14±6 years of different breed and gender, entered the study. For each horse, a caretaker was asked to fill in a validated temperament questionnaire. A novel object fear test (NOT), has been selected from literature to examine fearfulness. Temperature of the lacrimal caruncle was measured pre-test and post-test on 22 horses, representative of the whole sample. In order to assess discriminant validity of the NOT three human-animal relationship tests were performed on the same horses. Data were analyzed with descriptive, non-parametric and multivariate statistic methods. No significant differences were found between females and geldings for any of the measured variables. Horses that were described by caretakers as more prone to panic, vigilant, excitable, skittish and nervous ($p < 0.001$), needed significantly longer time to re-approach the novel object ($p < 0.01$). Eye temperature was significantly higher after the NOT compared to basal ($p < 0.01$), with subjects who did not re-approach the novel object tending to present larger increases ($p < 0.10$). Horses showing more fear related responses to the NOT did not show more negative reactions to humans during the human-animal tests. These results suggest that, to some extent, the NOT predicts horses' behaviour in real on-farm situations. Our findings reject the hypothesis that reactivity to humans and general fearfulness belong to the same basic feature of temperament. Importantly, infrared thermography proved to be useful in assessing physiological reactions of fear in horses.

Keywords: fear test, horse, infrared thermography, validity, welfare

1. Introduction

Fear in domestic animals has been defined by Boissy (1998) as a reaction of the perception of actual danger. Fear responses are characterized by behavioural and physiological modifications (Forkman et al., 2007): active defense (attack, menace), active flight (hiding, escape) and passive avoidance (freezing) are some of the behaviours that are frequently related to an underlying emotion of fear in animals (Erhard and Mendl, 1999). When experiencing fear, cardiovascular changes occur in different parts of the body with the ultimate effect of increasing perfusion pressure and redirecting blood follow to the Central Nervous System and skeletal musculature. Sport horses may be subject to different fearful events, for example, being transported and competing in different environments with novel stimuli and sounds (McGreevy and McLean, 2010), being approached by unfamiliar people or undergoing many handling and management procedures. Horses, as prey animals, have a tendency to escape from frightening stimuli and may show flight reactions which can be dangerous for both horse and man (Christensen et al., 2008, 2005; McGreevy and McLean, 2010): Keeling (1999) demonstrated that in equitation sports, many serious human injuries occur as a result of unexpected horse fear reactions. Because owners often misunderstand the reason for the development of such behaviours in their horses, attempts at correcting them often involve suppression or punishment based approaches (Hothersall and Casey, 2012). Although repeated subjugation of undesirable fear responses may ultimately appear to solve the overt behavioural reaction, this method can cause short- or long-term stress (McGreevy and McLean, 2010) and can worsen the problem or lead to the development of alternative avoidance strategies such as abnormal behaviours (Hothersall and Casey, 2012). Besides possible problems caused by inappropriate human reactions to fear displays, a long-term negative emotional state related to fear can *per se* cause chronic stress and reduced welfare (Dantzer and Mormede, 1983; Désiré et al., 2006; Minch et al., 2008; Willner et al., 1992).

Due to the aforementioned reasons, it is blatantly obvious that fear in horses plays an important role in their welfare, and thus it is important that it is recognized and assessed accordingly. Various fear tests have been used to determine temperament characteristics in horses: novel object (e.g. Anderson et al., 1999; Christensen et al., 2008, 2005; Seaman et al., 2002; Visser et al., 2003b, 2002; Wolff et al., 1997), novel arena (e.g. Le Scolan et al., 1997; Seaman et al., 2002; Wolff et al., 1997), and restraint and human fear tests (e.g. Le Scolan et al., 1997; Visser et al., 2003b, 2001; Wolff et al., 1997). The novel object test (NOT) is an experimental situation where the animal is exposed to an unknown stimulus to provoke a fear reaction. Although it is not possible to attribute a given measure to any single emotion, time to approach the new stimulus appears to be one of the most appropriate indicators of fearfulness (Górecka-Bruzda et al., 2011; Wolff et al., 1997). Feasibility under field conditions and ease and duration of fear tests are important characteristics for them to be applied as well as reliability and validity (Górecka-Bruzda et al., 2011). Validity means the degree to which a test measures what it purports to measure (Martin and Bateson, 1993; Weiblinger et al., 2006). Predictive validity measures the ability of an indicator to predict some later criterion (Cronbach and Meehl, 1955). In order to assess predictive validity of fear tests, different studies investigated their correlation with surveys via questionnaires which aimed to detect those characteristics of temperament in horses that influence their habitual behaviour (e.g. Anderson et al., 1999; Le Scolan et al., 1997; Momozawa et al., 2007, 2003; Morris et al., 2002a, 2002b). Respondents were generally caretakers or riding teachers who were familiar with horses, thus their responses were based on long-term observation and were not influenced by a temporary change in equine behaviour, which may occur in behavioural tests (Momozawa et al., 2005).

Discriminant validity analyzes the divergence between measures of conceptually unrelated concepts, for instance fear and human-animal relationship, and has seldom been evaluated for fear tests (Górecka-Bruzda et al., 2011; Visser et al., 2003b). Convergent validity regards the relationships between independent measures of the same conceptually related construct (Weiblinger et al., 2006). Assessment of convergent validity of fear tests usually considers whether their outcome is related to physiological changes due to fear. Some of the most frequently used physiological indicators are heart rate (e.g. Christensen et al., 2008; Momozawa et al., 2003), heart rate variability (e.g. Rietmann et al., 2004; Stewart et al., 2008c; Visser et al., 2002; von Borell et al., 2007), cortisol concentration (e.g. Anderson et al., 1999; Cook et al., 2001; Flauger et al., 2010; Stewart et al., 2008a), and Infrared Thermography (IRT). Infrared Thermography can be used to detect changes in peripheral blood flow (which causes changes in body heat) as a response to fear induced stress. Studies in different animal species have revealed that after a stressing event, the small areas around the posterior border of the eyelid and the caruncula lacrimalis change temperature. This area has rich capillary beds innervated by the sympathetic system (e.g. McGreevy et al., 2012; Stewart et al., 2009, 2007) and thus represents an ideal place for measuring local changes in blood flow resulting from tuning of the Autonomic Nervous System. Stewart et al. (2007) measured an increase in eye temperature in cows after intramuscular injection of ACTH, CRH and epinephrine. Research carried out on different species correlated increased eye temperature with cortisol concentrations in response to pain (Stewart et al., 2008b, 2008c), stress (Ludwig et al., 2007; Stewart et al., 2007; Valera et al., 2012), and fear (Stewart et al., 2008a). In a study on horses undergoing stressful situations Valera et al. (2012) found that the eye temperature increased as a consequence of stress. Similar results were found by Hall et al. (2001) who found a higher eye temperature in horses lunged with the Pessoa Training Aid (held responsible for increasing the stress during training) than horses without. Bartolomé et al. (2013), were able to demonstrate a correlation between an increase in heart rate and eye temperature after jumping competitions. Cook et al. (2001) investigated the underlying causes of increase in eye temperature in horses and found that it was correlated to activation of the HPA axis.

To our knowledge changes in superficial temperature during fear exposure have never been studied in horses. This study aims to assess the feasibility and predictive, convergent, and discriminant validity of a fear test in adult sport horses and investigates whether the exposure to a fearful stimulus induces a thermographic change in eye temperature.

2. Methods

This study was conducted in agreement with ISAE ethical guidelines (ISAE Ethics Committee, 2002) on adult non-pregnant horses and no animals underwent more than minimal distress. In addition, if horses displayed any hyper-reactive behaviour that could compromise the horse or the assessors' safety, the test was immediately ended and the observer left the box (this was recorded as a result).

2.1 Animals

Experiments took place from January to May 2013 at six different riding centers in Northern Italy. A total of 50 adult riding horses (mean age 14±6 years) of different sex (30 geldings, 16 mares, 4 stallions) were used in the study. Horse breeds were variously distributed and comprised warmblood horses, draft horses and thoroughbreds. All horses were stabled in single boxes with daily access to group paddocks for one to ten hours. Straw bedding was used in two centers, whereas horses were kept on wood shavings in the remaining three centers. Horses were fed three times a day with hay and concentrated industrial feed, depending on the type of activity they carried out. Water was provided ad libitum.

2.2 Questionnaire survey

Six caretakers (one per riding center) completed the questionnaires for the 50 tested horses; the number of questionnaires filled in per caretaker varied from six to ten. The questionnaire was developed and validated by Momozawa et al. (2005) and contained 20 questions regarding horse temperament (table 1). The responses were ranked on a scale from one to nine, with one being the lowest rank for each item. Two animal welfare experts translated the questionnaire into Italian; the mother tongue of both translators was Italian and their level of English was advanced. In a second round, the authors discussed and refined some of the items, which they felt might be difficult to interpret.

2.3 Behavioural tests

Four behaviour tests were chosen and are described in paragraphs 2.3.1 to 2.3.4. All tests were conducted on the same day and in the same housing conditions. Horses were tested at least one hour before work and between meals to avoid possible distractions and confounding food motivation. A map of the facility was drawn before testing the horses in order to facilitate the randomization of the testing order. To avoid habituation, horses kept in adjacent boxes were not tested consecutively. The test order was designed to firstly measure reactivity to a human, followed by the fear test. Two female experimenters (aged 24-28yrs), experienced in the field of animal welfare, conducted the tests. The first assessor performed the tests, while the second assessor scored the reactions of the horse to the different tests from a distance and without interfering with the test performance. To maintain consistency, the assessors always wore the same type and colour of clothing at all the riding centers, including appropriate safety clothing (e.g. accident prevention shoes) to reduce risk of injuries. Preventive safety measures always included making sure that there were no obvious physical hazards in the environment. Prior to the first assessment, both assessors familiarized themselves with the tests by researching relevant scientific literature and performing preliminary practical trials with a trainer familiar with the experimental procedures.

2.3.1 Fear Test (NOT)

For the fear test (NOT) an object which was not familiar to the horses was used. The procedure was derived and adapted from the work conducted by Górecka-Bruzda et al. (2011). A green, 1.5 l, plastic bottle, filled with small stones and attached by a 4 m cord, was placed at the box entrance and the cord was hung over the box door to keep the bottle at a height of approximately 1.5 m. In the original test the plastic container was placed next to the feeding bucket. The latency time to explore (sniffing, touching) the novel object was measured (first latency). When the horse approached, or after 300 seconds, the experimenter released the cord allowing the bottle to drop, thus emitting an unexpected, muffled noise. Latency to re-approach the bottle was then measured (second latency). The test was considered finished when the horse re-approached the bottle or after 300 seconds.

2.3.2 Avoidance Distance Test (AD)

At a distance of 2 m from the door of the horse-box, the observer waited until the horse's attention was directed towards them, and then slowly began to approach the horse at approximately one step per second. The observer never made direct eye contact with the horse; conversely they kept their eyes focused on the muzzle and an arm raised in front of them at an angle of 45°, with the palm facing downwards. The test terminated at any point when the horse showed an avoidance reaction (taking steps away from the observer or turning of the head). In such instances, a score of 0 was assigned. If the horse remained stationary and accepted being touched by the observer, a score of 1 was recorded.

2.3.3 Voluntary Animal Approach Test (VAA)

The assessor stood in front of the horse-box with their body at an angle of approximately 45°, and placed one hand on the box door whilst remaining motionless for 20 seconds. The latency until the horse approached and touched the hand was measured. If the horse did not approach the experimenter, a score of "more than 20 seconds" was given. The behaviour of the horse was also recorded on a three-point scale: 0 was given when the horse was aggressive (ears back, trying to kick, trying to bite, rearing); 1 when the horse showed no interest in human presence; 2 when the horse was interested and friendly (sniffing, turning the head toward the observer, approaching).

2.3.4 Forced Human Approach Test (FHA)

Once the horse had touched the experimenter or after a period of 20 seconds had passed with no signs of aggression shown, the assessor entered the box and approached the horse. Remaining approximately 0.5 m from the animal, the assessor placed a hand on the horse's neck and walked slowly to the rear of the horse maintaining contact with the horse. The behaviour toward the observer was recorder on a three-point scale: 0 was given when the horse did not allow the observer to touch them; 1 when the horse allowed the observer to touch them, but then tried to move away; 2 when the horse allowed the touch.

2.4 Infrared Thermography

On a group of subjects (N=22) from 3 riding centers and representative of the whole sample, eye temperature pre-test and post-test was evaluated. This group was composed of horses of different breed and gender (10 mares, four stallions and eight geldings), aged between three and 27 years (mean=13). An infrared camera (NEC AVIO TVS500, Nippon Avionics Co., LTD, Tokyo, Japan) with standard optic system was used to record the temperature (°C) of the lacrimal caruncle. The thermographic infrared images were captured by a certified technician (E.H.).

Lacrimal caruncle was chosen as target area based on information derived from (Bartolomé et al., 2013; Cook et al., 2001; McGreevy et al., 2012; Stewart et al., 2009) and because its temperature is not influenced by the presence of hair. In our study it was not possible to regulate room temperature and humidity but they were relatively stable across all situations (min=19.30 °C, max=21.00 °C, mean=19.73 °C).

To optimize the accuracy of the thermographic image and to reduce sources of noise, before every work session the same image of a Lambert surface was taken to define the radiance emission and to nullify the effect of surface reflections on tested animals (Mallick et al., 2005). Only images perfectly on focus were used. To determine the caruncle temperature, Grayess IRT analyzer 6.0

software (Grayess, 2007) was used and the maximum temperature (°C) within a circular area traced around the area was measured. This maximum value was used for subsequent analysis.

All the horses undergoing this procedure were accustomed to being restrained with head collar and a loose rope. In order to collect sharp images without using potentially stressful restraint methods, all the thermographic images were taken while the subject was gently restrained by holding the lead rope fixed to the head collar, allowing enough movement away from the approaching observer should the horse want to retreat. All horses were scanned from the same angle (90°) and distance (approximately 0.5 m) inside their own box. Five images were taken before and five images immediately after the test. All thermographic data was analyzed with Grayess-IRTAnalyzer (GRAYESS Inc., Bradenton, FL, USA) software.

2.5 Statistics

Data was entered into Microsoft Excel (Microsoft Corporation 2010) and then analyzed with SPSS statistical package (IBM SPSS Statistic 21). Descriptive statistics including relative proportions, minimum and maximum values, median, mean and standard deviations were calculated. The data was tested for normality using the Kolmogorov-Smirnov test. The U Mann-Withney test was used to verify if the gender of horses affected the questionnaire scores or the test outcomes. Differences were considered to be statistically significant if $p \leq 0.05$. Factor analysis was performed using the principle factor method for factor extraction, to evaluate any relationship between questionnaire items. A correlation matrix with varimax rotation was used, and factor scores were calculated for horses when the factor's Eigen value was greater than 1. A TwoStep Cluster analysis with automatic determination of the number of clusters was performed on questionnaire items relating to "fearfulness/anxiety" (as determined by Factor analysis) and outcomes of the NOT, in order to identify groups of horses that are similar to each other for the considered variables. The TwoStep clustering algorithm handles both continuous and categorical variables, continuous variables are z-standardized by default in order to make them comparable. The U Mann-Withney test was used to verify if the horses assigned to different clusters significantly differed for the considered variables. A match-paired Wilcoxon's test was used to compare thermographic data before and after the test and analysis of variance ANOVA was used to compare thermographic variations between horses who did or did not approach the novel object. The Kruskall Wallis ANOVA test was used to evaluate if the horses showing more intense fear reactions to the NOT also showed higher reactivity to the human-animal tests.

3. Results and discussion

The startling novel object test chosen as a reference (Górecka-Bruzda et al., 2011) and further refined in this study was selected because it is used in horses for measuring fear and its validity has been confirmed in a previous scientific work, although only in cold blood horses. It was also promising in terms of feasibility as it is of simple execution, it can be performed in the horse home box and its lead time is relatively short. However, prior to considering implementation in an on-farm welfare assessment protocol, refinement of the original test was deemed necessary to avoid possible conflicting motivations initially caused by proximity of the novel object to the food bucket. Our results revealed that the NOT was feasible under field conditions in sport horses. No safety issues were encountered, no tests had to be interrupted because of dangerous reactions of horses and all owners showed good acceptability of the procedure adopted to test the animals. Total time required to perform the test revealed substantial individual variability, ranging from 0 to 600 sec (mean 141±177 sec), mean latency time to first approach the bottle was 23±45 sec, and horses needed 27±34 sec to re-approach the bottle after it had been dropped in the box.

Table 2 reports the scores (min, max, median and standard deviation) of each questionnaire item. Horses were prevalently described by their caretakers as trainable, friendly towards people, with a good memory, cooperative, docile and were easy to get onto the trailer, as attested by high scores in these descriptors.

No significant differences were found in questionnaire scores or NOT results between females and geldings (U Mann-Whitney $p > 0.05$). Stallions were not compared to the other genders due to their limited number (N=4). Several authors investigated the effect of gender on personality traits of horses of different breeds and ages, using diverse methods and coming to different conclusions (Bartolomé et al., 2013; Kędzierski and Janczarek, 2009; Maros et al., 2010; Momozawa et al., 2007; Rietmann et al., 2004; Seaman et al., 2002; Visser et al., 2002; Wolff et al., 1997). Our results are consistent with Rietmann et al., (2004) who found that geldings did not differ from mares in any investigated measure of mental stress during training (HR, HRV and stress-related behaviour) and Seaman et al. (2002), who found no significant differences between the factor scores of mares and geldings subjected to three different behavioural tests (an arena test, an unknown person test and a novel object test). However, our findings are in contrast with studies by Momozawa et al. (2007) and Maros et al. (2010) who found differences between sexes in the response to a behavioural isolation test (Momozawa et al., 2007) and in the behaviour following a response to familiar humans (Maros et al., 2010). These dissimilarities between researches may be attributed to the diverse temperamental traits investigated using different experimental settings. Results of this study confirm that most of the differences between subjects seem to relate to individual behavioural differences and not to gender.

3.1 Predictive validity

In order to assess predictive validity of the NOT, the relation with a validated questionnaire (Momozawa et al., 2005) was investigated. Most results concerning predictive validity are similar to those obtained by Górecka-Bruzda et al. (2011) in cold blood horses. Table 3 shows the outcomes of the PCA performed on the scores of the questionnaire items. The analysis identified four main factors with Eigenvectors greater than 1, which together explain 61.9% of the variation between horses. Figure 1 represents the PCA loadings on the first two factors. The first factor, accounting for 31.3% of the total variance, shows high negative loadings for “nervousness”, “excitability”, “panic”, “inconsistent emotionality”, and “skittishness” suggesting that horses registering high negative scores on this factor can be described as more aroused and fearful than horses with high positive scores. These questionnaire items were considered for further analysis as the authors assumed that they could potentially be related to other indicators of fearfulness as the latency to approach a novel object. The first factor is also characterized by positive loadings of questionnaire items relating to trainability (“concentration”, “trainability”, “cooperation”, “perseverance”, “trailer”) and attitude toward humans (“docility”, “friendliness toward people”). Fearfulness, attitude toward humans and trainability might have common background in the sense that owners could have inappropriate reactions to fear displays affecting horses’ propensity to cooperate with humas. The second factor accounts for 13.6% of the total variance and shows high positive loadings for “memory” and “vigilance” as opposed to “friendliness toward horses”, suggesting that horses with high positive loadings on this factor tend to be more alert. The meaning of the other two factors, accounting for 10.2% and 6.6% of the total variance respectively, seems more elusive. The third factor shows high loadings for “curiosity” and “stubbornness” opposed to “timidity”. Only “competitiveness” belongs to factor four, so this factor retains the name “competitiveness”. The results of two questionnaire items -“stubbornness” and “friendliness toward horses”- are difficult to explain unambiguously as they appear not to be meaningfully associated with the others. One possible explanation is that the owners interpreted these questions differently. One possible problem with interpretation of the questionnaire is that it was developed for a specific population (Japanese) and respondents with a different cultural background might interpret it differently. To avoid these drawbacks, the questionnaire has been discussed among authors, as described in Method section. Despite these precautions, our results indicate that some questions could have been interpreted differently, hence correct wording of questionnaire items is essential.

A TwoStep Cluster analysis was performed on questionnaire items relating to "arousal/anxiety" (negative loadings on the first factor) and latency to approach and re-approach the bottle in the fear test, in order to identify groups of horses that are similar to each other for the considered variables. Two clusters were found based on the seven input variables selected. Fifty-two percent (N=26) of the horses were assigned to the first cluster and 48% (N=24) to the second. Horses in cluster 2 needed significantly more time to approach the bottle after it was dropped (U Mann-Whitney $p < 0.01$) and were described by their caretakers as more prone to panic, vigilant, excitable, skittish and nervous (U Mann-Whitney $p < 0.001$) (Figure 2). However, they did not differ in the latency time to approach the bottle when it was first placed at the box entrance (U Mann-Whitney $p > 0.05$). The bottle, when used as a static novel object, probably did not possess features that induced a clear reaction of fear enabling the differentiation of horses with various levels of fearfulness. Other studies revealed a moderate correlation between behaviour test outcomes and subjective evaluations of horse temperament provided by caretakers (Flentje, 2008; McCall et al., 2006; Visser et al., 2003b). For example, Momozawa et al. (2007, 2003) found comparable results in studies investigating correlations between the caretakers' responses about ordinary behaviours, heart rate, behaviour and latency times recorded during a Balloon Reaction Test or an isolation stress test. Although questionnaire surveys have the advantage of being based on long-term observation, they have the flaw of being subject to bias based on respondents' personal beliefs and temperament. Moreover, they should be carried out solely among those who are familiar with the behaviour of horses under different circumstances (Momozawa et al., 2007), as was the case in this study. When feasible and valid, standardised behaviour tests represent a preferable asset to people who deal with horse temperament evaluation in a broad range of facilities as they prevent unreliability of participants' responses. Relationships between results of the NOT and evaluation of caretakers suggest that, to some extent, the NOT outcomes represent a fearfulness temperamental trait.

3.2 Convergent validity

Convergent validity of the NOT was evaluated by examining relations between the test outcomes and variation of lacrimal caruncle temperature. This study shows for the first time that lacrimal caruncle temperature of horses undergoing the NOT was significantly higher after the test compared to basal (mean temperature before the test: 35.90±0.59 °C; mean temperature after the test: 36.19±0.60 °C; Wilcoxon's $p < 0.01$), indicating the presence of a physiological response to the test. Examples of thermographic pictures taken before and after the NOT are presented in Figure 3 (columns B and C, respectively). As shown in the figure, the temperature of the caruncle was higher in the post-test period (yellow and white areas), whereas it was relatively low before the NOT (orange areas). Also Nakayama et al. (2005) detected transient increases in temperature in the eye regions of four macachi resus (*Macaca mulatta*) during the stimulation of a potentially threatening person. Increased caruncle temperature was described by Stewart et al. (2007) in dairy cows injected with ACTH, CRH and epinephrine. Although, the same authors reported contradictory findings in cattle undergoing fear-eliciting (being hit with a plastic tube on the rump, being startled by the sudden waving of a plastic bag, restraint, electric prod, startled accompanied by shouting) (Stewart et al., 2008a) or painful stimuli (disbudding with or without local anesthetic) (Stewart et al., 2008b). A possible reason for discrepancy between these studies may be due to the nature of the fear stimuli used, as some of them might have caused pain besides fear. The magnitude of temperature variation was related to the intensity of reaction to the NOT: subjects who did not re-approach the bottle after it had been dropped in the box had a higher increase in lacrimal caruncle temperature (ANOVA $p < 0.1$) (Figure 4). These results confirm that horses who experienced intense negative emotions during the fear test presented more evident behavioural signs related to fear (they do not re-approach the bottle) and higher variation in lacrimal caruncle temperature. Analogously to Vianna and Carrive (2005), who investigated changes in laboratory rats undergoing a conditioned fear response to footshock chambers and who found that tail temperature was sensitive to the level of arousal, the findings of the present study suggest that the stronger the arousal, the stronger the physiological response.

3.3 Discriminant validity

Discriminant validity of the NOT was studied by examining the possible relationship with fear of people. Table 4 shows descriptive results of the three human-animal relationship tests. Fifty-six percent of the horses did not show any avoidance behaviour when approached by the assessor in the AD test. In VAA and FHA tests, only 6.1 % of the horses displayed negative reactions. The horses which had shown avoidance reactions during the AD test or negative reactions to the FHA test did not need more time to re-approach the novel object compared to horses that had expressed an amicable behaviour towards humans during human-animal relationship tests (ANOVA Kruskal-Wallis $p > 0.05$).

These results suggest that fear reactions shown in the NOT are not related to the responses of horses towards unfamiliar humans. Other research failed to prove that different behaviour tests effectively distinguish between fear of people and a more general fearfulness trait (Górecka-Bruzda et al., 2011). In this study, similarly to Visser et al. (2003a), we demonstrated that the NOT is specifically informative of the general fearfulness trait. These results do not support the hypothesis that reactivity to humans and general fearfulness belong to the same basic feature of temperament.

5. Conclusion and future directions

The fear test originally developed by Górecka-Bruzda et al. (2011), refined and adapted by the authors of this study to horses of different breeds and to different conditions, proved to be a valid measure of general fearfulness of horses and could be easily implemented for use in an on-farm welfare assessment protocol. The relatively limited number of subjects on which the thermographic measures were performed (N=22) constitutes a limiting factor for the generalization of the results of the present study. In any case, our results are a valid indication for a relationship between superficial eye temperature and fear emotion. This study provides a new angle on mechanisms regulating interaction between horse emotions and behaviour. Future studies should consider a larger sample of horses in order to substantiate the results and to measure time to return to baseline eye temperature after the fear stimulus.

Acknowledgment

The authors would like to thank the EU VII Framework program (FP7-KBBE-2010-4) for financing the Animal Welfare Indicators (AWIN) project. We are grateful to the riding school owners who allowed us to assess their horses. A special thanks goes to Dr Francesco Tozzi, from the ASL of Varese, who patiently helped us in recruiting horse facilities. The authors would like to thank Leigh Anne Margaret Murray and Kirk Ford for their extensive and professional revisions of language and structure.

Ethic Statement

The work described in this article has been carried out in accordance with EU Directive 2010/63/EU for animal experiments. Horses were involved in this study at the request of their owner on a voluntary basis and approval by an Ethical Committee was not required.

Conflict of interests statement

The Authors certify that there is not any actual or potential conflict of interest.

Authorship statement

The idea for the paper was conceived by Dai, Minero, Canali

The experiments were designed by Dai, Cogi, Minero

The experiments were performed by Dai, Cogi, Heinzl

The data were analyzed by Dai and Minero

The paper was written by Dai, Cogi, Heinzl, Dalla Costa, Canali, Minero

References

Anderson, M., Friend, T., Evans, J., Burshong, D., 1999. Behavioral assessment of horses in therapeutic riding programs. Appl. Anim. Behav. Sci. 63, 11–24.

Bartolomé, E., Sánchez, M., Molina, A., Schaefer, A., Cervantes, I., Valera, M., 2013. Using eye temperature and heart rate for stress assessment in young horses competing in jumping competitions and its possible influence on sport performance. Animal 7, 2044–53. doi:10.1017/S1751731113001626

Boissy, A., 1998. Fear and fearfulness in determining behavior, in: Grandin, T. (Ed.), Genetics and the Behaviour of Domestic Animals. Academic Press, San Diego, USA, pp. 67–111.

Christensen, J.W., Keeling, L.J., Nielsen, B.L., 2005. Responses of horses to novel visual, olfactory and auditory stimuli. Appl. Anim. Behav. Sci. 93, 53–65. doi:10.1016/j.applanim.2005.06.017

Christensen, J.W., Malmkvist, J., Nielsen, B.L., Keeling, L.J., 2008. Effects of a calm companion on fear reactions in naive test horses. Equine Vet. J. 40, 46–50. doi:10.2746/042516408X245171

Cook, N., Schaefer, A., Warren, L., Burwash, L., Anderson, M., Baron, V., 2001. Adrenocortical and metabolic responses to ACTH injection in horses: an assessment by salivary cortisol and infrared thermography of the eye. Can. J. Anim. Sci. 81, 621.

Cronbach, L.J., Meehl, P.E., 1955. Construct validity for psycological tests. Psycol. Bull. 52, 283–302.

Dantzer, R., Mormede, P., 1983. Stress in farm animals - a need for reevaluation. J. Anim. Sci. 57, 6–18.

Désiré, L., Veissier, I., Despres, G., Delval, E., Toporenko, G., Boissy, A., 2006. Appraisal process in sheep (Ovis aries): interactive effect of suddenness and unfamiliarity on cardiac and behavioral responses. J. Comp. Psychol. 120, 280–287.

Erhard, H., Mendl, M., 1999. Tonic immobility and emergence time in pigs - more evidence for behavioural strategies. Appl. Anim. Behav. Sci. 61, 227–237.

Flauger, B., Krueger, K., Gerhards, H., Möstl, E., 2010. Simplified method to measure glucocorticoid metabolites in faeces of horses. Vet. Res. Commun. 34, 185–95. doi:10.1007/s11259-010-9344-y

Flentje, R., 2008. How reliable are standardised behaviour tests and are they valid in predicting the suitability for use in police horses? MSc Dissertation, University of Liverpool.

Forkman, B., Boissy, A., Meunier-Salaün, M., Canali, E., Jones, R., 2007. A critical review of fear tests used on cattle, pigs, sheep, poultry and horses. Physiol. Behav. 92, 340–374. doi:10.1016/j.physbeh.2007.03.016

Górecka-Bruzda, A., Jastrzębska, E., Sosnowska, Z., Jaworski, Z., Jezierski, T., Chruszczewski, M.H., 2011. Reactivity to humans and fearfulness tests: Field validation in Polish Cold Blood Horses. Appl. Anim. Behav. Sci. 133, 207–215. doi:10.1016/j.applanim.2011.05.011

Grayess, 2007. IRT analyser users' manual.

Hall, C., Burton, K., Maycock, E., Wragg, E., 2001. A preliminary study into the use of infrared thermography as a means of assessing the horse's response to different training methods, in: Hartmann, E., Blokhuis, M., Fransson, C., Dalin, G. (Eds.), 6th International Equitation Science Symposium. Uppsala, Sweden, p. 64.

Hothersall, B., Casey, R., 2012. Undesired behaviour in horses: A review of their development, prevention, management and association with welfare. Equine Vet. Educ. 24, 479–485. doi:10.1111/j.2042-3292.2011.00296.x

ISAE Ethics Committee, 2002. Ethical Treatment of Animals in Applied Animal Behaviour Research. URL http://www.applied-ethology.org/ethical_guidelines.html (accessed 4.14.14).

Kędzierski, W., Janczarek, I., 2009. Sex-related effect of early training on stress in young trotters as expressed by heart rate. Anim. Sci. Pap. Reports 27, 23–32.

Keeling, L., Blomberg, A., Ladewig, J., 1999. Horse-riding accidents: when the human–animal relationship goes wrong!, in: 33rd International Congress of the International Society for Applied Ethology. Lillehammer, Norway, p. 86.

Le Scolan, N., Hausberger, M., Wolff, A., 1997. Stability over situations in temperamental traits of horses as revealed by experimental and scoring approaches. Behav. Process. 41, 209–221.

Ludwig, N., Gargano, M., Luzi, F., Carenzi, C., Verga, M., 2007. Technical note: Applicability of infrared thermography as a non invasive measurement of stress in rabbit. World Rabbit Sci. 15, 199–205.

Mallick, S.P., Zickler, T.E., Kriegman, D.J., Belhumeur, P.N., 2005. Beyond Lambert: Reconstructing Specular Surfaces Using Color, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Ieee, pp. 619–626. doi:10.1109/CVPR.2005.88

Maros, K., Boross, B., Kubinyi, E., 2010. Approach and follow behaviour – possible indicators of the human – horse relationship. Interact. Stud. 11, 410–427. doi:10.1075/is.11.3.05mar

Martin, P., Bateson, P., 1993. Measuring Behaviour: An Introductory Guide. Cambridge University Press, Cambridge.

McCall, C., Hall, S., McElhenney, W., Cummins, K., 2006. Evaluation and comparison of four methods of ranking horses based on reactivity. Appl. Anim. Behav. Sci. 96, 115–127. doi:10.1016/j.applanim.2005.04.021

McGreevy, P., McLean, A., 2010. Equitation science. Wiley-Blackwell, Chichester, West Sussex, UK.

McGreevy, P., Warren-Smith, A., Guisard, Y., 2012. The effect of double bridles and jaw-clamping crank nosebands on temperature of eyes and facial skin of horses. J. Vet. Behav. Clin. Appl. Res. 7, 142–148. doi:10.1016/j.jveb.2011.08.001

Minch, H., Berghaus, R., Harvey, S., Reeves, D., Crowell-Davis, S.L., 2008. A novel method for lifting weanling research pigs. J. Vet. Behav. Clin. Appl. Res. 3, 266–275.

Momozawa, Y., Kusunose, R., Kikusui, T., Takeuchi, Y., Mori, Y., 2005. Assessment of equine temperament questionnaire by comparing factor structure between two separate surveys. Appl. Anim. Behav. Sci. 92, 77–84. doi:10.1016/j.applanim.2004.11.006

Momozawa, Y., Ono, T., Sato, F., Kikusui, T., Takeuchi, Y., Mori, Y., Kusunose, R., 2003. Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behavior test. Appl. Anim. Behav. Sci. 84, 127–138. doi:10.1016/j.applanim.2003.08.001

Momozawa, Y., Terada, M., Sato, F., Kikusui, T., Takeuchi, Y., Kusunose, R., Mori, Y., 2007. Assessing equine anxiety-related parameters using an isolation test in combination with a questionnaire survey. J. Vet. Med. Sci. 69, 945–50.

Morris, P.H., Gale, A., Duffy, K., 2002a. Can judges agree on the personality of horses? Pers. Individ. Dif. 33, 67–81. doi:10.1016/S0191-8869(01)00136-2

Morris, P.H., Gale, A., Howe, S., 2002b. The factor structure of horse personality. Anthrozoos A Multidiscip. J. Interact. People Anim. 15, 300–322.

Nakayama, K., Goto, S., Kuraoka, K., Nakamura, K., 2005. Decrease in nasal temperature of rhesus monkeys (Macaca mulatta) in negative emotional state. Physiol. Behav. 84, 783–90. doi:10.1016/j.physbeh.2005.03.009

Rietmann, T., Stuart, A., Bernasconi, P., Stauffacher, M., Auer, J., Weishaupt, M., 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. doi:10.1016/j.applanim.2004.02.016

Seaman, S., Davidson, H., Waran, N., 2002. How reliable is temperament assessment in the domestic horse (Equus caballus)? Appl. Anim. Behav. Sci. 78, 175–191.

Stewart, M., Schaefer, A.L., Haley, D.B., Colyn, J., Cook, N.J., Stafford, K.J., Webster, J.R., 2008a. Infrared thermography as a non-invasive method for detecting fear-related responses of cattle to handling procedures. Anim. Welf. 17, 387–393.

Stewart, M., Stafford, K.J., Dowling, S.K., Schaefer, a L., Webster, J.R., 2008b. Eye temperature and heart rate variability of calves disbudded with or without local anaesthetic. Physiol. Behav. 93, 789–97. doi:10.1016/j.physbeh.2007.11.044

Stewart, M., Stookey, J., Stafford, K., Tucker, C., Rogers, A., Dowling, S., Verkerk, G., Schaefer, A., Webster, J., 2009. Effects of local anesthetic and a nonsteroidal antiinflammatory drug on pain responses of dairy calves to hot-iron dehorning. J. Dairy Sci. 92, 1512–9. doi:10.3168/jds.2008-1578

Stewart, M., Webster, J., Schaefer, A., Stafford, K., 2008c. Infrared thermography and heart rate variability for non-invasive assessment of animal welfare. ANZCCART News 21, 1–4.

Stewart, M., Webster, J., Verkerk, G., Schaefer, A., Colyn, J., Stafford, K., 2007. Non-invasive measurement of stress in dairy cows using infrared thermography. Physiol. Behav. 92, 520–5. doi:10.1016/j.physbeh.2007.04.034

Valera, M., Bartolomé, E., Sánchez, M.J., Molina, A., Cook, N., Schaefer, A., 2012. Changes in Eye Temperature and Stress Assessment in Horses During Show Jumping Competitions. J. Equine Vet. Sci. 32, 827–830. doi:10.1016/j.jevs.2012.03.005

Vianna, D.D.M.L., Carrive, P., 2005. Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. Eur. J. Neurosci. 21, 2505–2512. doi:10.1111/j.1460-9568.2005.04073.x

Visser, E.K., Van Reenen, C.G., Engel, B., Schilder, M.B.H., Barneveld, A., Blokhuis, H.J., 2003a. The association between performance in show-jumping and personality traits earlier in life. Appl. Anim. Behav. Sci. 82, 279–295. doi:10.1016/S0168-1591(03)00083-2

Visser, E.K., van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. Appl. Anim. Behav. Sci. 74, 241–258. doi:10.1016/S0168-1591(01)00177-0

Visser, E.K., van Reenen, C.G., Rundgren, M., Zatterqvist, M., Morgan, K., Blokhuis, H.J., 2003b. Responses of horses in behavioural tests correlate with temperament assessed by riders. Equine Vet. J. 35, 176–183.

Visser, E.K., van Reenen, C.G., van der Werf, J.T.N., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2002. Heart rate and heart rate variability during a novel object test and a handling test in young horses. Physiol. Behav. 76, 289–96.

Von Borell, E., Langbein, J., Després, G., Hansen, S., Leterrier, C., Marchant-Forde, J., Marchant-Forde, R., Minero, M., Mohr, E., Prunier, A., Valance, D., Veissier, I., 2007. Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals - A review. Physiol. Behav. 92, 293–316. doi:10.1016/j.physbeh.2007.01.007

Weiblinger, S., Boivin, X., Pedersen, V., Tosi, M.V., Janczak, A.M., Visser, E.K., Jones, R.B., 2006. Assessing the human-animal relationship in farmed species: a critical review. Appl. Anim. Behav. Sci. 101, 185–242.

Willner, P., Muscat, R., Papp, M., 1992. Chronic mild stress-induced anhedonia: a realistic animal model of depression. Neurosci. Biobehav. Rev. 16, 525–534.

Wolff, A., Hausberger, M., Le Scolan, N., 1997. Experimental tests to assess emotionality in horses. Behav. Processes 40, 209–221. doi:10.1016/S0376-6357(97)00784-5