Applied Animal Behaviour Science 290 (2025) 106716

Contents lists available at ScienceDirect

Applied Animal Behaviour Science

journal homepage: www.elsevier.com/locate/applanim

The ability to interpret affective states in horses' body language is associated with experience with animals

Moritz Nicolai Braun [a,*,1], Sina Bülow [b,2], Alicia Müller-Klein [a], Ulrike Link-Dorner [c], Tanja Michael [a,3], Johanna Lass-Hennemann [a,4]

[a] Division of Clinical Psychology and Psychotherapy, Department of Psychology, Saarland University, Saarbrücken 66123, Germany

[b] Department of Psychology, Christian-Albrechts-Universität zu Kiel, Kiel 24118, Germany

[c] Institut für Tiergestützte Ausbildung und Therapie (ITAT), Eschringerstraße 70, Saarbrücken 66131, Germany

A R T I C L E I N F O

Keywords:
Emotion
Equine
Body language
Interspecies-communication
Replication

A B S T R A C T

In light of an increasing interest in the human ability to read equine emotions, Braun et al. (2024) recently showed that horse-experienced individuals performed better in interpreting affective states in horses' body language than horse-inexperienced individuals. Further, individuals with a high emotion recognition ability performed better in interpreting horses' affective states than individuals with a low emotion recognition ability. The authors operationalized the ability to interpret affective states from horses' body language by means of the behavior identification test (BIT). In the BIT, participants are presented with 32 photographs of horses that express different affective states and, for each photograph, are asked to choose which out of eight affective states the depicted horse expresses in its body language. In the present study, we scrutinized the replicability and the robustness of these findings. Further, we tested the idea that a better ability to interpret affective states in horses' body language is associated with experience with animals in general but even more so with horses in particular. We did so by employing the design by Braun et al. (2024) in a novel sample, expanded by a number of quasi-experimental variables (i.e., animal/horse contact while growing-up; current pet/horse ownership; contact with animals/horses during work). We directly replicated the effect reported by Braun et al. (2024) that horse-experienced individuals outperform horse-inexperienced individuals in the BIT. Further, we demonstrated that a higher emotion recognition ability was associated with a better ability to interpret affective states in horses' body language. Moreover, we found strong evidence in support of the notion that a better ability to interpret affective states in horses' body language is associated with experience with animals in general and, more strongly, with horses in particular, be it during childhood or currently. We call future research to test the idea that the ability to interpret affective states in horses' body language can be trained through contact with horses or other animals by means of experimental designs to identify a potential causal relation.

1. Introduction

In recent years, there has been an increasing interest in the human ability to read equine emotions. While some studies investigated the human ability to interpret horses' vocalizations (Greenall et al., 2022; Merkies et al., 2022) others focused on the human ability to identify affective states in horses' body language (Bell et al., 2019; Braun et al., 2024; Fox, 2023; Gronqvist et al., 2017; Guinnefollau et al., 2019). The latter suggest an influence of previous experience with horses. For instance, fourth year veterinary students outperformed first year students in correctly interpreting equine behavior from a photograph (Guinnefollau et al., 2019). Moreover, experience of horse ownership helped with the recognition of horses' negative affective states visible in short videos (Bell et al., 2019). More recently, Braun et al. (2024) asked

* Correspondence to: Saarland University, Department of Psychology, Division of Clinical Psychology and Psychotherapy, Campus A1 3, Saarbrücken 66123, Germany.

E-mail addresses: moritz.braun@uni-saarland.de, moritz.nicolai.braun@gmail.com (M.N. Braun).

1 ORCID iD: https://orcid.org/0000-0002-9779-0293.
2 ORCID iD: https://orcid.org/0009-0007-6263-7540.
3 ORCID iD: https://orcid.org/0000-0002-2409-3817.
4 ORCID iD: https://orcid.org/0000-0001-7278-2521.

https://doi.org/10.1016/j.applanim.2025.106716
Received 26 March 2025; Received in revised form 1 June 2025; Accepted 17 June 2025
Available online 18 June 2025

their participants to interpret affective states expressed in the body language of photographed horses. Precisely, for each of the 32 photographs presented, the participants had to choose which out of eight affective states the depicted horse expresses in its body language. Crucially, before completing this Behavior Identification Test (BIT), the participants were asked to indicate their previous experience with horses and to complete the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001; Oakley et al., 2016) as a measure of their emotion recognition ability. In the BIT, participants performed far from perfect, with an average of 12 (out of 32) correctly identified affective states. This suggests that humans are not particularly good in reading equine emotions or that horses are not very good at expressing their affective states to humans – prerequisites, as we argue, for effective communication and bonding between humans and horses and thus, for safe equine activities. However, an analysis of variance (ANOVA) yielded that horse-experienced participants significantly outperformed horse-inexperienced participants in the BIT. Moreover, participants with high emotion recognition ability significantly outperformed participants with low emotion recognition ability.

Especially in light of the replication crisis (e.g., Baker, 2016; Pashler and Wagenmakers, 2012; Shrout and Rodgers, 2018) and the associated view of replication as the scientific gold standard (e.g., Jasny et al., 2011; Simons, 2014), we deem it important to scrutinize the replicability and robustness of the findings from Braun et al. (2024). In the present study, we therefore asked a novel sample of participants to complete the BIT. As in the original study, we additionally asked our participants to indicate their previous experience with horses and to complete the RMET. To scrutinize the replicability of the original findings, we tested the following hypotheses:

(H1). More experience with horses is associated with a better ability to interpret affective states in horses' body language.

(H2). A better emotion recognition ability is associated with a better ability to interpret affective states in horses' body language.

To scrutinize the robustness of the original findings, we tested (H1) and (H2) by means of both the same analytical method as in the original study (i.e. an ANOVA) and a different analytical method (i.e. correlations).

In addition, we also tested the idea that more experience with animals in general is associated with a better ability to interpret affective states in horses' body language. To do so, we investigated the influence of a number of quasi-experimental variables on the participants' BIT scores. In detail, we asked participants to report (1) whether their work involves living animals and if so, whether their work involves horses, (2) whether they grew up with animals and if so, to list the animals they grew up with, and (3) whether they have pets or other animals that they regularly spend time with and if so, to list their pets and the animals they regularly spend time with. Using the BIT and these variables, we tested the hypotheses that participants whose work involves living animals outperform participants whose work does not (see Greenall et al., 2022 for some evidence in this direction), that participants who grew up with animals outperform participants who did not, and that participants who have pets or other animals that they regularly spend time with outperform participants who do not. We deem it likely that experience with horses is significantly stronger associated with a better ability to interpret affective states in horses' body language than experience with other animals. Thus, we also tested the hypotheses that participants whose work involves horses outperform participants whose work involves other animals, that participants who grew up with horses outperform participants who grew up with other animals, and that participants who have a horse or regularly spend time with a horse outperform participants who have a pet other than a horse or regularly spend time with an animal other than a horse.

Further, we took an exploratory look at whether interest in modern equestrian sport as practiced at the Olympic Games and the attitude towards this sport, whether training horses with a form of positive reinforcement (e.g., food and a clicker or a comparable form, such as a word as a marker signal or scratching as a reward), and whether the type of humans' diet influence the ability to interpret affective states in horses' body language.

2. Methods

This study was pre-registered with (H1) and (H2) under AsPredicted (https://aspredicted.org/357r-8xqz.pdf). All data and code can be found in the Open Science Framework project associated with this study (https://osf.io/3d9nx/). This study was approved by our university's ethics committee (reference number 17-12).

2.1. Sample

We collected data from $n = 216$ participants (159 women, 54 men, 3 diverse with a $Mean_{Age}$ of 33.85 years, $SD_{Age} = 14.29$ years, range: 18–75 years). Suitable for inclusion were German-speaking adults who did not take part in the study by Braun et al. (2024). A priori sample size calculation using G*Power (Version 3.1.9.7; Faul et al., 2009, 2007; yielded a sample size of $n = 170$ participants to detect the $\eta^2_G = .05$ (the smallest effect found in the complete sample by Braun et al., 2024) equivalent of $\rho = 0.22$ with a power of 90 % at an α-level of 0.05 in a correlation analysis. With the final sample size of $n = 216$, our study has a power greater than 95 % to detect said effect.

2.2. Materials and measures

General Questionnaire. First, we asked the participants to answer questions regarding their gender, age, and type of diet (omnivore, vegetarian or vegan). We then asked the participants whether their work involves living animals and if so, to state which profession they work in and whether their work involves horses. Next, we asked the participants whether they grew up with animals and if so, to list the animals they grew up with, whether they have pets or other animals that they regularly spend time with and if so, to list their pets and the animals they regularly spend time with. Importantly, we then asked the participants about the frequency of their contact with horses. As in the study by Braun et al. (2024), the participants were asked to choose between 1) On a regular basis, 2) Every now and then, 3) Solely during childhood, 4) Passively (e.g., via kids you bring to a riding lesson), and 5) Never. Subsequently, we asked the participants whether they are interested in modern equestrian sport as practiced at the Olympic Games and about their attitude towards this sport (like, dislike or no opinion). Finally, we asked participants, whether they train horses with food and a clicker or a comparable form of positive reinforcement, such as a word as a marker signal or scratching as a reward (yes, no I don't use food or a clicker or anything like that or no I do not train).

RMET. As in the study by Braun et al. (2024), we employed the revised version of the "Reading the Mind in the Eyes" Test (RMET; Baron-Cohen et al., 2001; Oakley et al., 2016) to assess participants emotion recognition ability. We randomly and successively presented 36 photographs that the depict human eye sections and asked participants to choose, which of four adjectives (e.g. interested, joking, affectionate, contended; varying from photograph to photograph) best describes what the person in each of the photographs is feeling or thinking. The participants could score between 0 and 36 points in the RMET (correct answers are scored 1, incorrect answers 0). The higher the score, the better the participant's emotion recognition ability.

BIT. As in the study by Braun et al. (2024), participants were asked to complete the behavior identification test (BIT). In the BIT, 32 photographs of horses (the same photographs that were used in the study by Braun et al., 2024) that express different affective states (11 photographs of negative valence, 11 photographs of positive valence, and 10 photographs of neutral valence) are presented one after another in randomized order (for details regarding how these photographs were created and how the affective states were induced, see Braun et al., 2024). For each photograph, we asked the participants to choose the affective state that they assumed behind the horses' body language out of the following list: 1) Defense, Threat, Frustration, 2) Fear, Panic, 3) Attention, Positive Tension, 4) Relaxation, Exploration, 5) Happiness, 6) Nervousness, Negative Tension, 7) Neutral (Neither depicts a positive nor a negative emotion or state), 8) Pout, Being Offended. Importantly, options 5) and 8) served as distractors to reduce the guessing probability and are not depicted in any of the 32 photographs (see Braun et al., 2024). The participants could score between 0 and 32 points in the BIT (correct answers are scored 1, incorrect answers 0). The higher the score, the better the participant's ability to interpret affective states in horses' body language.

2.3. Procedure

We conducted the study online using LimeSurvy (LimeSurvey GmbH (n.d.)) and advertised it via social media and at our universities. Participants participated via QR code or link on their own devices. Each participant actively gave their informed consent to participate in our study, after having received general information about the procedure of the study and information about the anonymity of data. After completion of the general questionnaire, participants completed the RMET, followed by the BIT.

2.4. Statistical analyses

We aimed to scrutinize the robustness of the findings reported by Braun et al. (2024) by means of different analytical methods. Thus, we tested the hypothesis that more experience with horses is associated with a better ability to interpret affective states in horses' body language (H1), by computing a Spearman rank correlation between the participants' experience with horses and participants' BIT score with a one-sided test of significance. Further, we tested the hypothesis that a better emotion recognition ability is associated with a better ability to interpret affective states in horses' body language (H2) by computing a Pearson correlation between participants' RMET score and participants' BIT score with a one-sided test of significance.

In addition, we also analyzed our data using the approach originally employed by Braun et al. (2024), i.e., by computing a two-factorial ANOVA with the overall score in the BIT as the dependent variable and the two two-levelled factors "Experience" (Experienced vs. Inexperienced) and "Emotion Recognition Ability" (High vs. Low). Horse-experienced participants are defined as participants who reported to have contact with horses on a regular basis. All other participants are summarized as horse-inexperienced. This resulted in 66 participants (61 women, 4 men, 1 diverse) classified as experienced and 150 participants (98 women, 50 men, 2 diverse) classified as inexperienced. To create the two levels of the factor "Emotion Recognition Ability", we split the RMET overall score at the median.

We analyzed the influence of the quasi-experimental variables covered in the general questionnaire on the BIT score using independent samples *t*-tests and one-way ANOVAs, depending on whether the variable had two or three levels, respectively. As there are no published findings supporting our hypotheses regarding most of these variables, we ran all independent samples *t*-tests as two-sided to be more conservative.

We analyzed our data using R (Version 4.3.1, R Core Team, 2023) and R Studio (Version 2023.09.0; RStudio Team, 2023) and applied the significance criterion of $p < .05$ throughout analyses. Please see the analysis script in the Open Science Framework project associated with this study for the specific packages we used.

3. Results

3.1. Replication of Braun et al. (2024)

The descriptive statistics reported in Table 1 show that the performances in the BIT and the RMET were comparable with the performances reported by Braun et al. (2024). In support of H1, the Spearman rank correlation with a one-sided test of significance yielded that more experience with horses was associated with a better ability to interpret affective states in horses' body language ($\rho(214) = .56$, $p < .001$), indicating a large effect. In support of H2, the Pearson correlation with a one-sided test of significance yielded that a better emotion recognition ability was associated with a better ability to interpret affective states in horses' body language ($r(214) = .22$, $p < .001$), indicating a small effect.

H1 was also supported by the results from the ANOVA. The ANOVA yielded a significant main effect for Experience ($F(1, 212) = 97.49$, $p < .001$, $\eta^2_G = .31$), indicating a large effect, with higher BIT scores for experienced ($M = 15.65$, $SD = 3.51$) than for inexperienced participants ($M = 10.63$, $SD = 3.45$). However, the results from the ANOVA did not support H2. Although descriptively, participants with a high emotion recognition ability ($M = 12.50$, $SD = 4.08$) performed better in the BIT than participants with a low emotion recognition ability ($M = 11.86$, $SD = 4.24$), the main effect for Emotion Recognition Ability did not reach significance ($F(1, 212) = 2.83$, $p = .094$). The two-way interaction for Experience × Emotion Recognition Ability did also not reach significance ($F(1, 212) = 0.82$, $p = .366$).

3.2. Experience with animals in general and horses in particular

Participants who grew up with an animal ($M = 12.49$, $SD = 4.15$) performed better in the BIT ($t(214) = 2.29$, $p = .023$, $d = 0.38$), indicating a small effect, than participants who did not grow up with an animal ($M = 10.91$, $SD = 4.03$). Out of those participants who grew up with an animal, those who grew up with a horse ($M = 14.74$, $SD = 3.94$) performed better in the BIT ($t(169) = 5.84$, $p < .001$, $d = 0.93$), indicating a large effect, than those who grew up with another animal ($M = 11.21$, $SD = 3.71$). Similarly, participants who currently own a pet or currently have regular contact with an animal ($M = 13.60$, $SD = 4.06$) performed better in the BIT ($t(214) = 5.93$, $p < .001$, $d = 0.81$), indicating a large effect, than participants who do not ($M = 10.46$, $SD = 3.63$). Out of those participants who currently own a pet or currently have regular contact with an animal, those who currently own a horse or currently have regular contact with a horse ($M = 15.49$, $SD = 3.57$) performed better in the BIT ($t(115) = 6.58$, $p < .001$, $d = 1.23$), indicating a large effect, than those who do not ($M = 11.23$, $SD = 3.36$). Further, participants whose work involves living animals ($M = 14.60$, $SD = 3.73$) performed better in the BIT ($t(214) = 3.18$, $p = .002$, $d = 0.68$), indicating a moderate effect, than participants whose work does not ($M = 11.84$, $SD = 4.12$). As only five participants indicated that their work involves living animals but not horses, we could not further analyze whether working without horses but other living animals makes a difference. Note that the pattern of results does not change if the Bonferroni corrected significance criterion of $p < .025$ is applied for all *t*-tests (which could be argued for as necessary).

Table 1
BIT and RMET descriptive statistics in the present study and the study by Braun et al. (2024).

	Mean	Median	SD	Min	Max
BIT	12.16	12	4.17	4	23
BIT (Braun et al., 2024)	12.02	12	4.23	3	22
RMET	25.29	25	4.13	9	34
RMET (Braun et al., 2024)	24.40	25	3.82	15	35

Note. SD = Standard deviation. Min = Minimum. Max = Maximum. BIT = Behavior Identification Test. RMET = Reading the Mind in the Eyes Test.

3.3. Exploratory analyses

Participants who stated that they are interested in modern equestrian sport as practiced at the Olympic Games ($M = 14.59$, $SD = 4.35$) performed better in the BIT ($t(214) = 4.30$, $p < .001$, $d = 0.75$), indicating a moderate effect, than participants who stated that they are not interested ($M = 11.59$, $SD = 3.92$). A one-way ANOVA ($F(213, 2) = 10.56$, $p < .001$, $\eta^2_G = .09$), indicating a moderate effect, with post-hoc Tukey Honest Significant Differences revealed that participants who reported to have no opinion on modern equestrian sport ($M = 10.85$, $SD = 3.68$) scored lower in the BIT than participants who reported to like said sport ($M = 13.30$, $SD = 4.52$, $p = .014$) and participants who reported to dislike said sport ($M = 13.37$, $SD = 4.17$, $p < .001$). The latter two groups of participants did not differ significantly ($p = .996$). A one-way ANOVA ($F(213, 2) = 31.89$, $p < .001$, $\eta^2_G = .23$), indicating a large effect, with post-hoc Tukey Honest Significant Differences revealed that participants who reported to not train horses ($M = 10.88$, $SD = 3.68$) scored lower in the BIT than participants who reported to train horses with food and a clicker or a comparable form of positive reinforcement, such as a word as a marker signal or scratching as a reward ($M = 14.75$, $SD = 3.46$, $p < .001$) and participants who reported to not use food or a clicker or anything like that during training ($M = 16.57$, $SD = 4.29$, $p < .001$). The latter two groups of participants did not differ significantly ($p = .228$). A one-way ANOVA did not yield a significant influence of participants' type of diet on the BIT score ($F(213, 2) = 0.12$, $p = .884$).

4. Discussion

In light of an increasing interest in investigating the human ability to read equine emotions, Braun et al. (2024) recently showed that horse-experienced individuals performed better in interpreting affective states in horses' body language than horse-inexperienced individuals. Further, they showed that individuals with a high emotion recognition ability performed better in interpreting horses' affective states than individuals with a low emotion recognition ability. The authors operationalized the ability to interpret affective states from horses' body language by means of the behavior identification test (BIT). In the BIT, participants are presented 32 photographs of horses that express different affective states and, for each photograph, are asked to choose which out of eight affective states the depicted horse expresses in its body language.

In the present pre-registered and high-powered study, we replicated Braun et al.'s first finding by means of a different (i.e. correlation) and the same analytical method as in the original study (i.e. an ANOVA). Precisely, we found (1) that more experience with horses is strongly associated with a better ability to interpret affective states in horses' body language, indicated by a large correlation of experience with horses and BIT score, and (2) that horse-experienced individuals outperform horse-inexperienced individuals in the BIT. This supports the robustness of the findings by Braun et al. (2024). In addition, performances in the BIT were comparable with the performances of the participants in the study by Braun et al. (2024). These results strengthen the confidence in the effect reported by Braun et al. (2024) and are in line with previous research that suggests an influence of previous experience with horses (e.g., Bell et al., 2019; Fox, 2023; Gronqvist et al., 2017; Guinnefollau et al., 2019).

In support of the idea that a better ability to interpret affective states in horses' body language is associated with more experience with animals in general, we found that participants who grew up with an animal performed better in the BIT than participants who did not. Further, we found that participants who currently own a pet or currently have regular contact with an animal performed better in the BIT than participants who do not. In addition, we found that participants whose work involves living animals performed better in the BIT than participants whose work does not. In support of the idea that a better ability to interpret affective states in horses' body language is more strongly associated with experience with horses than experience with other animals, we found that out of those participants who grew up with an animal, those who grew up with a horse performed better in the BIT than those who grew up with another animal. Further, we found that out of those participants who currently own a pet or currently have regular contact with an animal, those who currently own a horse or currently have regular contact with a horse performed better in the than those who do not. As only five participants indicated that their work involves living animals but not horses, we could not further analyze whether working without horses but other living animals makes a difference. In sum, our results strongly support the notion that a better ability to interpret affective states in horses' body language is associated with experience with animals in general and, more strongly, with horses in particular, be it during childhood or currently. Future research could test the idea that the ability to interpret affective states in horses' body language can be trained through contact with horses or other animals to identify a potential causal relation. This could be done by employing experimental designs in which participants complete the BIT before and after some period of regular contact with horses or other animals, potentially enriched with information about horses' emotions and their expressions.

In the present study, we also found that a better emotion recognition ability was associated with a better ability to interpret affective states in horses' body language. This supports the finding by Braun et al. (2024) that participants with a high emotion recognition ability performed better in interpreting horses' affective states than participants with a low emotion recognition ability. However, we could not directly replicate this effect using the same analytical approach as Braun et al. (2024). There is a number of possible explanations for our failure to directly replicate this effect. One possible explanation is that the effect found by Braun et al. (2024) is not real. Another possible explanation is that by chance one out of ten times the ANOVA yields a non-significant result even if the effect exists and is of the size reported by Braun et al. (2024). In this case, the ANOVA we computed has a power of around 90 % to detect the effect of the size reported in the original study at an α-level of 0.05. In case the effect reported by Braun et al. (2024) was in fact of a smaller size, the chance of a non-significant result in the ANOVA would increase. Importantly, the results from the present study alone do not allow to decide which of the explanations holds true. However, given that we found support for this effect using the pre-registered correlation analysis and that descriptively, participants with a high emotion recognition ability performed better in the BIT than participants with a low emotion recognition ability, we think that the effect should not be dismissed but scrutinized further in future studies.

Our exploratory analyses yielded that participants with an interest in modern equestrian sports as practiced at the Olympic Games performed better in the BIT than participants without such an interest, while participants' opinion about said sport (like vs. dislike) did not make a difference. Further, our exploratory analyses yielded no significant difference between training forms (positive reinforcement vs. no positive reinforcement) and no significant influence of participants' type of diet.

4.1. Limitations

Before closing, we would like to note some limitations of our study. First, to be able to directly replicate the findings by Braun et al. (2024), we decided to stick to the design and, thus, the stimulus material used in the original study (i.e., the 32 photographs of the BIT). With the BIT, we aimed to develop a standardized set of photographs that (1) depict horses' affective states that were experimentally induced in a controlled environment and that (2) do not depict any contextual information that might influence the participants' judgements (as this might lead to misinterpretations, see Bell et al., 2019). To validate our photographs beyond the experimental induction, we further relied on expert judgements (Braun et al., 2024). However, while we believe that our experimental induction (which took place against the background of the depicted horses' previous learning experience) reliably creates photographs depicting horses' affective states, we have no additional data beyond the experts' judgements to proof this. Thus, our results should be interpreted with caution. Moreover, non-verbal communication is a complex dynamic process that is informed (and sometimes misguided, see Bell et al., 2019) by contextual information. Thus, relying on videos instead of photographs as the stimulus material might provide more comprehensive information regarding horses' affective states and might increase participants' performance by enabling them to incorporate contextual information and vocalizations. In light of the ongoing debate about valid postural indicators of positive affect in horses (Lesimple, 2020; Waran and Randle, 2017), the validity of the BIT may be enhanced by measuring additional indicators (e.g., physiological parameters). In sum, future research could employ the BIT and video material as well as incorporate other indicators such as physiological parameters to further scrutinize our findings. Second, as in the original study, the majority of our participants self-identified as women. To increase generalizability, future research could aim for gender-balanced samples.

5. Conclusion

In the present pre-registered and high-powered study, we directly replicated the effect by Braun et al. (2024) that horse-experienced individuals outperform horse-inexperienced individuals in the BIT. We further found that a better emotion recognition ability was associated with a better ability to interpret affective states in horses' body language as well as strong support for the notion that a better ability to interpret affective states in horses' body language is associated with experience with animals in general and, more strongly, with horses in particular. We call future research to test the idea that the ability to interpret affective states in horses' body language can be trained through contact with horses or other animals by means of experimental designs to identify a potential causal relation.

CRediT authorship contribution statement

Moritz Nicolai Braun: Writing – review & editing, Writing – original draft, Supervision, Project administration, Formal analysis, Data curation, Conceptualization. **Johanna Lass-Hennemann:** Writing – review & editing, Writing – original draft, Supervision, Resources. **Tanja Michael:** Writing – review & editing. **Sina Bülow:** Writing – review & editing, Project administration, Investigation, Conceptualization. **Ulrike Link-Dorner:** Writing – review & editing. **Alicia Müller-Klein:** Writing – review & editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Baker, M., 2016. Reproducibility crisis. Nature 533 (26), 353–366. https://doi.org/10.1038/533452a.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001. The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J. Child Psychol. Psychiatry 42 (2), 241–251. https://doi.org/10.1017/S0021963001006643.

Bell, C., Rogers, S., Taylor, J., Busby, D., 2019. Improving the recognition of equine affective states. Animals 9 (12), 1124.

Braun, M.N., Müller-Klein, A., Sopp, M.R., Michael, T., Link-Dorner, U., Lass-Hennemann, J., 2024. The human ability to interpret affective states in horses' body language: the role of emotion recognition ability and previous experience with horses. Appl. Anim. Behav. Sci. 271, 106171. https://doi.org/10.1016/j.applanim.2024.106171.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191. https://doi.org/10.3758/BF03193146.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.G., 2009. Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41 (4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Fox, S.A., 2023. Accuracy of Horse Affect Assessments: A Comparison of Equine Assisted Mental Health Professionals, Non-equine Assisted Mental Health Professionals, and Laypeople (Master thesis). University of Lethbridge, Faculty of Education.

Greenall, J.S., Cornu, L., Maigrot, A.L., De La Torre, M.P., Briefer, E.F., 2022. Age, empathy, familiarity, domestication and call features enhance human perception of animal emotion expressions. R. Soc. Open Sci. 9 (12), 221138. https://doi.org/10.1098/rsos.221138.

Gronqvist, G., Rogers, C., Gee, E., Martinez, A., Bolwell, C., 2017. Veterinary and equine science students' interpretation of horse behaviour. Animals 7 (8), 63. https://doi.org/10.3390/ani7080063.

Guinnefollau, L., Gee, E.K., Bolwell, C.F., Norman, E.J., Rogers, C.W., 2019. Benefits of animal exposure on veterinary students' understanding of equine behaviour and self-assessed equine handling skills. Animals 9 (9), 620. https://doi.org/10.3390/ani9090620.

Jasny, B.R., Chin, G., Chong, L., Vignieri, S., 2011. Again, and again, and again…. Science 334 (6060), 1225. https://doi.org/10.1126/science.334.6060.1225.

Lesimple, C., 2020. Indicators of horse welfare: state-of-the-art. Animals 10 (2), 294. https://doi.org/10.3390/ani10020294.

LimeSurvey GmbH, n.d. LimeSurvey: An Open Source Survey Tool. LimeSurvey GmbH. ⟨https://www.limesurvey.org⟩.

Merkies, K., Crouchman, E., Belliveau, H., 2022. Human ability to determine affective states in domestic horse whinnies. Anthrozoös 35 (3), 483–494. https://doi.org/10.1080/08927936.2021.1999605.

Oakley, B.F., Brewer, R., Bird, G., Catmur, C., 2016. Theory of mind is not theory of emotion: a cautionary note on the Reading the mind in the eyes test. J. Abnorm. Psychol. 125 (6), 818–823. https://doi.org/10.1037/abn0000182.

Pashler, H., Wagenmakers, E.J., 2012. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? Perspect. Psychol. Sci. 7 (6), 528–530. https://doi.org/10.1177/174569161246525.

R Core Team, 2023. R: A Language and Environment for Statistical Computing [Software]. R Foundation for Statistical Computing, Vienna, Austria. ⟨https://www.R-project.org/⟩.

RStudio Team, 2023. RStudio: Integrated Development for R [Software]. RStudio, PBC, Boston, MA. ⟨http://www.rstudio.com/⟩.

Shrout, P.E., Rodgers, J.L., 2018. Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. Annu. Rev. Psychol. 69 (1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845.

Simons, D.J., 2014. The value of direct replication. Perspect. Psychol. Sci. 9 (1), 76–80. https://doi.org/10.1177/174569161351475.

Waran, N., Randle, H., 2017. What we can measure, we can manage: the importance of using robust welfare indicators in equitation science. Appl. Anim. Behav. Sci. 190, 74–81. https://doi.org/10.1016/j.applanim.2017.02.016.