2023-03-02

Do you see what I see? Testing horses' ability to recognise real-life objects from 2D computer projections

Kappel, Sarah

https://pearl.plymouth.ac.uk/handle/10026.1/21003

10.1007/s10071-023-01761-6
Animal Cognition
Springer Science and Business Media LLC

Do you see what I see? Testing horses' ability to recognise real-life objects from 2D computer projections

Sarah Kappel[1], Marco A Ramirez Montes De Oca[2], Sarah Collins[1], Katherine Herborn[1], Mike Mendl[2] and Carole Fureix[1,2]

[1]University of Plymouth, School of Biological and Marine Sciences, Portland Square, Drake Circus, Plymouth, PL4 8AA, UK

[2]University of Bristol, Bristol Veterinary School, Langford house, BS40 5DU Langford, UK

Corresponding author: Sarah Kappel sarah.kappel@plymouth.ac.uk

Abstract

The use of 2-dimensional representations (*e.g.* photographs or digital images) of real-life physical objects has been an important tool in studies of animal cognition. Horses are reported to recognise objects and individuals (conspecifics and humans) from printed photographs, but it is unclear whether image recognition is also true for digital images, *e.g.* computer projections. We expected that horses trained to discriminate between two real-life objects would show the same learnt response to digital images of these objects indicating that the images were perceived as objects, or representations of such. Riding-school horses (N=27) learnt to touch one of two objects (target object counterbalanced between horses) to instantly receive a food reward. After discrimination learning (three consecutive sessions of 8/10 correct trials), horses were immediately tested with on-screen images of the objects over 10 image trials interspersed with five real object trials. At first image presentation, all but two horses spontaneously responded to the images with the learnt behaviour by contacting one of the two images, but the number of horses touching the correct image was not different from chance (14/27 horses, $p>0.05$). Only one horse touched the correct image above chance level across 10 image trials. (9/10 correct responses, $p = 0.021$). Our findings thus question whether horses recognise real-life objects from digital images. We discuss how methodological factors and individual differences (*i.e.* age, welfare state) might have influenced animals' response to the images, and the importance of validating the suitability of stimuli of this kind for cognitive studies in horses.

Keywords: image recognition, horse cognition, individual cognitive performance, equines

1. Introduction

Visual 2-dimensional representations (*e.g.* printed photos, digital images, silhouettes, videos) are used as substitutes for real-life objects, or individuals, in cognition studies of non-human animals, including horses. Screen-displayed visuals are of advantage in research as stimulus timing and presentation of identical stimuli can be repeatedly presented to the same or to different subject animals (D'Eath 1998). However, scientific evidence of object-image recognition in animals is not always consistent (reviewed in Fagot 2000; Bovet and Vauclair 2000; Weisman and Spetch 2010). This might be because pictures designed for the human eye may not result in the same sensory experiences in other species with different functional visual systems (Fagot and Parron 2010; Weisman and Spetch 2010). Moreover, how images are perceived and cognitively processed is not fully understood for most animal species (Fagot 2000; Fagot et al. 2010). For instance, Fagot et al. (2010) proposed that animals could 'read' images using different processing modes. In a mode of *confusion*, images and their real-life exemplars are perceived and treated as functionally and physically the same thing. Conversely, in a mode of *independence*, images could be perceived as different from their referents without making an association between objects and their images. In a processing mode of *equivalence*, images are understood as representations of their referents (*i.e.* images are used as referential cues for real-life objects, Fagot 2000; Fagot et al. 2010).

A variety of factors, including cognitive limitations or experience with images, could influence which processing mode is deployed by animals and ultimately lead to differences in how images are treated by humans and other animal species (Fagot and Parron 2010). Therefore, the suitability of artificial representations (*e.g.* digital images, videos) for animal studies is likely to depend on the purpose of the stimuli. For instance, if images are used to imitate real stimuli in behavioural experiments, animals need to respond to images in a comparable way to how they respond to real stimuli (D'Eath 1998).

Investigating image recognition is challenging because pictures can never be identical to their 3D referents given the lack of dimensionality, depth cues and olfactory characteristics, which results in substantial sensory differences between objects and their 2D imitations (Bovet and Vauclair 2000; Aust and Huber 2006). Prior to image processing, the perceptual abilities of the viewer also need to be considered, for instance, whether an animal is able to identify an object from an image despite the lack of depth cues or additional cues (*e.g.* reflectance of photographic surface, Fagot and Parron 2010).

Unlike in humans, the visual field of horses is mainly monocular (*i.e.* visual input is received from just one eye, (Waring 2003). Binocular vision allowing depth perception is only possible within a relative small area in front of the horses' head (55°- 65°; Hughes 1977) extending downwards along the midsagittal plane (the vertical axis dividing the head in left/right) at approximately 75°, enabling horses to view the ground in front of them with both eyes (Duke-Elder 1958). A blind spot interrupts the almost panoramic visual field in front of the horses' forehead (Waring 2003). In addition, visual acuity is much poorer in horses compared to most other terrestrial mammals (Rørvang et al. 2020). Horses have dichromatic vision resulting in similar colour perception to humans affected by red-green blindness (Hanggi et al. 2007). However, equine vision is highly adapted to low-light conditions with a high ratio of rods to cones and a reflecting tapetum lucidum enabling scotopic vision (*i.e.* ability to see under low light conditions) superior to that of humans (Hanggi and Ingersoll 2009a). Given these visual differences, it appears that humans and horses see the world differently (Saslow 2002). This raises the question of whether artificial stimuli such as digital images generated through computer projections are suitable representations of real-life objects for horses and other ungulate species sharing these traits (e.g. cattle, goats, sheep; Jacobs et al. 1998). Hence, further validation whether horses recognise the content of digital stimuli is necessary.

Generally, two different experimental approaches are applied to test image recognition in animals (reviewed in Bovet and Vauclair 2000; Weisman and Spetch 2010). For one, animals' spontaneous responses to artificial representations of biologically relevant stimuli (*e.g.* photos of food, prey, predator or conspecifics) is tested as an indication of direct transfer (*i.e.* images are treated as the same as objects). In this case, the same adaptive behaviour is provoked by the artificial representations as if the real referent was present (Bovet and Vauclair 2000; Weisman and Spetch 2010). A study in sheep, another ungulate species, found that animals respond to the image of a sheep with species-specific social behaviour (*e.g.* sniffing of the anogenital region and the head) and the sheep image appears to have fear-reducing effects on socially isolated sheep comparable to the presence of real conspecifics (Vandenheede and Bouissou 1994). Interestingly, a human image did not result in the same fear response as elicited by a real human, suggesting that different stimuli types may be processed differently by sheep (*i.e.* sheep image possibly confused with a real sheep whereas the human images was not treated as a substitute; Vandenheede and Bouissou 1994). Horses also respond to 2D and 3D horse imitations (photograph, life-size model) with sniffing behaviour near the head and flank areas corresponding to their natural approach of conspecifics, while an incomplete horse drawing and a dog image were not approached (Grzimek 1943). These observations might suggest that horses are able to recognise conspecifics based on specific cues, such as social cues conveyed by a near-realistic 3D model and photograph but not a drawing. However, approach and sniffing behaviours are also associated with exploration meaning that using explorative responses as outcome measures is not specific to image recognition alone and could result from other motivations, such as gathering novel information. Similar reasoning may apply to other studies that use spontaneous approach behaviours to indicate image recognition in horses (*e.g.* Smith et al. 2016; Wathan et al. 2016). Physiological changes (mean heart rate) measured alongside horse behaviour were interpreted by the authors as support for horses' ability to differentiate between emotional stimuli, although cross-validation through multiple physiological measure (*e.g.* HRV indices to infer autonomic response; von Borell et al. 2007) could have strengthened these findings even more.

An alternative to the above-described adaptive behaviour responses is studying animals' ability to transfer acquired (operant) responses associated with real life objects to their pictorial representations (Bovet and Vauclair 2000). For example, Cabe (1976) trained pigeons to discriminate between two solid objects (one rectangular block and a cross) by pecking the rewarded stimulus. The birds spontaneously transferred the learnt discrimination rule when the objects were replaced by pictorial representations (*e.g.* black-and-white photographs, white-on-black silhouettes) demonstrating that pigeons are able to recognise objects from images (Cabe 1976). Using a similar approach, Hanggi (2001) reported that, after multiple presentations, horses (N=2) were able to transfer a learnt behaviour (contact object with nose for food) from real objects (various toys varying in colour, shape and size) to their pictures, indicating image recognition. However, the ability to categorise images does not automatically provide evidence of representational insight (*i.e.* the subject understands what the image stands for; Aust and Huber 2006). The horses might have learnt to discriminate between the images during repeated testing, *e.g.* based on invariant features between images (*e.g.* colour, shapes, or distribution of light/dark patterns) unrelated to the real objects. According to the author, this explanation seems unlikely given the large number and diversity of objects tested (Hanggi 2001). However, the same two horses were previously reported to understand shared characteristics between stimuli (pattern rules; (Hanggi 1999), indicating their ability of categorisation learning, which one animal was reported to still remember several years later (Hanggi and Ingersoll 2009b).

Experimental biases and ambiguity of outcome measures can further hamper the validity of image recognition evidence. For instance, it has been reported that horses can recognise humans from images because they were not only able to differentiate between happy and angry human faces, but also appear to possess emotional memory (Proops et al. 2018). Horses were described as reacting "appropriately" following the theory of emotional lateralisation (*i.e.* left-eye bias for humans with angry faces and more time engaging in stress-related displacement behaviours) when encountering the real human hours after they had seen a photo of the same person displaying an angry face. Another study suggested that horses have the ability to cross-modally recognise the emotional states of familiar caretakers and stranger when presented with on-screen image of human faces and voice recordings (Nakamura et al. 2018). However, due to experimental limitations (*e.g.* in Proops et al. (2018), horses were kept in different conditions between tests, non-specificity of response behaviours (*e.g.* scratching, floor sniffing; these activities that are also expressed in other contexts (Waring 2003)) and statistical weakness (*e.g.* no control conditions), the robustness of these findings has been questioned (Amici 2019). Moreover, inferring evidence of recognition from emotional responses might not be straightforward in absence of control (*i.e.* non-emotional) comparisons. Hence, it is possible that the horses' response could have been associated with image-inherited cues unrelated to the emotional image content (*e.g.* image colours, brightness or contrast). The study by Lansade et al. (2020a) reduced experimental biases by training horses first to reliably select a screen image showing one of four human faces instead of images of objects (novel objects differing on each trial), thereby priming horses to respond to content-specific information. The horses significantly discriminated between the familiar faces and a novel face. When a photo of the horses' keeper replaced the training faces, the animals again selected the keeper image at above chance level suggesting that the keepers' faces were also identified as familiar. Alternatively, the keeper images might have been more similar to each of the training images than the novel images. In a follow-up study using on-screen images, Lansade et al. (2020b) controlled for this and found that horses can reliably select familiar faces paired against unfamiliar faces, despite removing photo colour, external cues (hairstyle), or facial features (eyes).

Overall, given a variety of experimental difficulties in this area, there is still a need for further evidence of the ability of horses to recognise the content of screen images and their relationship with real-life objects. The motivation of this study was therefore to test if horses spontaneously respond to digital images of two real-life objects, which they had previously learnt to discriminate. We predicted that horses would touch the images of the correct (rewarded) object at a level above chance if they recognised the images as real objects or representations of such. We only tested horses' transfer ability from real-life objects to on-screen images, and not the reverse (*i.e.* training horses with images to test discrimination with their real-life counterparts), to gain evidence that digital images are suitable stimuli for cognitive tests in this species. For this, we developed relative simple and practical testing approach. For the same reason, we only used two real-life objects.

Animals' performance in cognitive tests can be influenced by individual characteristics, including personality (Carere and Locurto 2011; Dougherty and Guillette 2018), learning speed, and motivation to engage in the task (reviewed in Rowe and Healy 2014). In horses, age (Krueger et al. 2014), sex (Murphy et al. 2004), but also emotional state (Christensen et al. 2012; Valenchon et al. 2013), and welfare status (reviewed in Hausberger et al. 2019) have been identified as sources of individual variation in cognitive performance. Therefore, we tested each horse in a total of 10 trials and assessed the effects of intrinsic (*i.e.* age, welfare score) and experimental factors (*e.g.* type of target, trial order, facility) on horses' performance.

2. Methods

2.1. Ethical statement

This study was approved by the Animal Welfare and Ethical Review Body of the University of Plymouth (ETHICS-41-2020). The experimental procedure complied with the UK Animals (Scientific Procedures) Act 1986 (ASPA) and followed the ARRIVE guidelines 2.0. The horses belonged to two UK riding schools who consented the use of their animals. Housing, care and health check was provided by the riding schools. The animals remained at their home facility at the end of the study, except one horse that was relocated during our data collection for reasons not related to this study. Horses that did not learn the object discrimination in stage 1 were excluded from the object recognition test in stage 2.

2.2. Animals and housing

In total, 36 horses of mixed breeds from two UK riding schools (yard A: N=17, mean ± SD age 10.6 ± 2.5 years; yard B: N=19, 16.6 ± 6.5 years, of which three animals did not complete training at this yard as one was relocated and two became aggressive towards nearby conspecifics during training) were trained in an object discrimination test (ODT, stage 1). All horses that completed stage 1 (*i.e.* discrimination between the real objects; N=28) were tested in the on-screen object recognition test (ORT, stage 2). However, one horse was scared of the test setup and was therefore excluded from testing, resulting in a total of 27 horses (16 from yard A of which 6 were females, 11 from yard B of which 4 were females) used in the ORT. The horses were used in riding lessons approx. 3-7h per week. In both facilities, horses were kept in single stalls, or tie-stalls, with full, or limited visual/physical contact to conspecifics during daytime (details of horses in Supplementary Information, Table 1). All horses had pasture access (in stable groups) at night and/or during parts of the day. Hay provision was restricted (*i.e.* facilities adjusted hay allowance based on body weight), and horses received an additional adjusted diet (at yard B, brand Thunderbrook Equestrian), or not (at yard A where horses were "on a diet" due to the lowered workload associated with COVID-19 restrictions). Water was freely accessible through automatic troughs in yard A and provided with water buckets in yard B.

2.3. Experimental design

The experimental design consisted of two stages summarised in Fig. 1. In stage 1, the horses were trained to discriminate between two real objects by touching the rewarded (target) object with their muzzle in order to receive a food reward before their spontaneous response to on-screen images was tested in stage 2.

2.3.1. Object discrimination – stage 1

All horses were first trained inside their stall by a single familiar person (experimenter SK) to respond to the real objects and discriminate between the target (rewarded) and an unrewarded object. The horses were able to move around freely (although six horses at yard B were tethered as they were kept in tie-stalls). Two objects (*kong™*: red dog toy, Ø 10cm, length 16cm; *ring*: doughnut-shaped dog toy, Ø 20cm, depth 4cm, with dark and light blue stripes, see Fig. 1) used as target objects were mounted onto a 50 cm wooden stick to facilitate the presentation of the objects in different positions and at distance to the experimenter. Which object a horse received as *target* (rewarded object) was pseudo-randomly allocated, ensuring that the numbers of horses trained with the same target was evenly distributed across yards. As only horses that completed ODT and learnt the discrimination within the 5 training sessions were used in ORT, the final number of horses tested in ORT with the ring and kong™ as target object was 11 and 16, respectively.

The first training step consisted of shaping horses' **response to the target object** using instrumental conditioning. The experimenter moved toward the horses' shoulder (whichever side that was most accessible) hiding the target behind her back. Standing at the shoulder height, she then slowly moved the object into view for the horse and held the target at approx. 20-30cm from the horses' muzzle (approx. 1.0 – 1.2m above the ground depending on horses' height). The horse could voluntarily move towards the object and contact with the object was never forced. Upon the first voluntary contact, the horse was instantly rewarded with a piece of carrot retrieved from a treat bag attached to the experimenter's waist at her back. At the same time, the target was moved behind the experimenter's back. Within 5s of rewarding the horse, the same motion of moving the target near the horses' muzzle was repeated and the horse was instantly rewarded upon voluntary contact. All contacts with the object only (regardless of where on the object and with which part of the muzzle) were rewarded. The target training was repeated for 10 consecutive trials. The experimenter then left the stall to refill the treat bag again with 10 pieces of carrots and repeated this training step so that each horse received a total of 20 single target trials.

After a 2-min break, 10 single target trials were conducted again to remind the horses of the correct (familiar) target before a second unfamiliar object was introduced. The experimenter followed the same procedure as before to present the objects, except that now two objects were shown to the horses simultaneously for object

discrimination training (**ODT,** see Fig.1A). For this, the experimenter moved both objects simultaneously from behind her back to in front of the horses' head holding each object by its handle in one hand at approx. 1.0-1.2m above the ground and with objects separated approx. 0.4-0.6m. If horses touched the unrewarded object, the objects were shortly moved behind the experimenter's back for 5s time-out before starting a new trial. If the unrewarded object was consecutively touched over three trials, the experimenter only presented the target to the horse (to remind it of the target, and guarantee that the horse received a reward and maintained motivation). The number of these *forced* trials was not recorded as this occurred rarely. If a horse did not touch any objects within 30s, this response was regarded as incorrect, and a new trial was started. On each trial, the experimenter slightly altered her position relative to the horse, in which location and side, from the horses' perspective, the objects were shown, and alternated the hand used to reward the horse. These changes were done to avoid the horses develop side biases, or learning by association which object to contact relative to the handler (*e.g.* always chose object in experimenter's left hand). In addition, the side of object presentation was pseudo-randomly selected by the experimenter with the same object never being presented on the same side more than twice during consecutively. Depending on horse availability, each horse received a maximum of two ODT training sessions per day, each comprising four trial blocks and 10 discrimination trials per block with 2min breaks between each block. Horses were trained over a maximum of five sessions (equal to 200 discrimination trials in total), and with a maximum of three days between sessions. Training of three horses at yard 2 was interrupted due to COVID-19 restrictions and resumed 6 months later starting from ODT. For these horses, only trials conducted after the break were included in the data analysis. Learning criterion (LC) required to move to stage 2 (testing) was defined as performing eight or more correct responses per trial block over three consecutive trial blocks. The eight horses that did not reach LC within five training sessions were not tested in stage 2.

2.3.2. Object recognition test - stage 2

Stage 2 consisted of the on-screen object **recognition** test (**ORT**) and was divided into three steps (see Fig.1B). Pre-tests conducted in the horses' stall (step 1) and the test arena (step 2) using the real objects serving as verification of reliable discrimination performance before the horses were tested with images in the screen test (step 3).

2.3.2.1. Pre-test in stall

The horses first received 10 single target trials conducted by the experimenter in the horses' home stalls. A second unfamiliar handler (MR) then entered the horses' stall alongside the experimenter to take hold of the lead rope, hence mimicking the handler's presence later in the test stage. The handler stood next to the horses' left shoulder, with his back turned to the horse and wearing noise-cancelling headphones to remain blinded to which of the two objects was the target. The experimenter presented the two objects for 10 trials as done in the ODT, except that the objects were now always presented in front of the horses' head at approx. 1–1.5m height, *i.e.* at similar position as to where the images replacing the real object would later occur in the screen test. The handler's role was to reward the horse as indicated by the experimenter (saying her name to indicate an incorrect response, or the handlers' name to indicate a correct response) whilst remaining blind to the correct target to avoid any conscious or unconscious signalling from the handler (*i.e.* 'clever Hans effects', Pfungst and Rahn 1911) during later stages of testing. If the horse performed ≥8 correct responses out of 10 in the pre-test, it was immediately taken to the test area for the screen test. Horses that did not perform as such were re-tested in the same manner after a break (of varied duration for practical reasons, *e.g.* horse availability).

2.3.2.2. Pre-screen test (PST)

The horse was led into the test area (familiar indoor riding arena) where a back-projection polyvinyl chloride screen (1.6m W x 2.5m H) was set up. A multi-coloured pole (normally used as training item and familiar to the horse) serving as visual marker was placed on the ground directly in front of the screen at approx. 50cm distance to indicate the position of the horse during testing. The horse was habituated to the screen (first turned off, then turned on not showing any images) and test equipment until it stood calmly in front of the screen. The screen was then turned off again and the handler positioned himself approx. 1m away from the ground pole by the horses' left shoulder, turning his back towards the screen (position allowing him to stay blind to the images to be shown in the next phase). The experimenter stood in front of the horse (between the ground pole and screen) towards the right side of its head. She retrieved the real objects from a bucket and conducted 10 ODT trials following the same procedure as during the pre-test in stall (*i.e.* the experimenter presented the objects and indicated to the blinded handler when to give deliver the reward). This was done to test if the horse still discriminated between the real objects in this different context (arena rather than stall). After five trials, the experiment briefly moved behind the screen (out of view from the horse[1]) to habituate the horse to her movement and absence. After five seconds, she returned to her original position in front of the screen and conducted five more trials.

If the horse performed ≥8 correct responses out of 10, the experimenter stepped behind the screen to start the screen test. If the horse performed below this level, it was led around the arena for approx. 2min and the pre-screen test was repeated. In total, horses received a maximum of six pre-screen tests, with a maximum of three daily (number derived from pilot observations where one horse needed six pre-screen tests to move to the screen test). All horses performed at the required criterion within six pre-screen tests.

2.3.2.3. Screen test

In preparation for the screen test, each object was photographed three times using a Fujifilm X-T100 digital camera (focal lens 23mm). Images were edited to remove the background so that only the object and wooden handle were visible in the final images (see Fig.1). Three versions of computer presentations (Microsoft PowerPoint) were created, each consisting of 10 stimulus slides. Each slide contained one image of each object side-by side on white background. Within the three presentations, the location of target images was balanced (50% left) and pseudo-randomised so that the target object was shown no more than twice in a row on the same side. The order and side of images varied between the three presentations to control for order effects. Additionally, the images were randomly rotated around their horizontal plane to change the position of the wooden handle. Later on screen, the images were shown approx. 1.1-1.2m above the ground and at 0.5-0.7m distance from each other.

Each stimulus slide was preceded by a white blank slide, except for the slides prior to stimulus slides 4, 7 and 9, which were black, indicating the points in the test at which real object trials were to be conducted (later described). Each horse was tested with only one out of the three presentations (equally spread across tested horses). Which presentation was projected was unknown to the experimenter at the time of testing, ensuring blindness to the target location (since the only slide she saw when starting playing the presentation was a blank slide).

The screen test started immediately following the pre-screen test. The images were broadcasted from a laptop (Lenovo ThinkPad 13) *via* a LCD-projector (HITACHI CP-WX3030WN) placed at approx. 2.5m distance behind the screen. Standing next to the laptop, the experimenter used a remote control to start the slide show and advance the slides (thereby moving as little as possible to avoid any distracting noise). The first slide was blank but the experimenter advanced to the first stimulus slide as soon as the horses' head was straight in front of the screen

---

[1] but the experimenter could observe the horse *via* a web cam connected to a computer (Lenovo ideapad) serving as monitor

(monitored *via* the web cam allowing to see the horse and the screen content). As soon as the horse contacted one of the images, the stimulus slide was immediately advanced to the next blank slide. At the same time, the experimenter indicated to the blind handler whether a reward should be delivered. A trial commenced as soon as the horses' head was straight in front of the screen again resulting in variable inter-trial intervals. The stimulus slides advanced automatically to the next blank slide after 20s if no contact was made. In case the horse moved away from the screen immediately after trial onset (approx. within <2s after stimulus onset), the presentation was moved to the previous blank slide and the trial repeated as soon as the horses' head was back in a straight position in front of the screen.

In total, 10 trials with images were conducted, interspersed with real object trials (where the experimenter returned to her position by the horse). Two object trials were conducted after image trial 3 and 6, and one object trial was conducted after image trial 8 (*i.e.* five objects trials in total conducted during the screen test). The real object trials were conducted as per the pre-screen test procedure, to remind the horses of the properties of the real objects, and to test whether they were still motivated to touch the objects, even if the images were not touched. To avoid that horses learnt to respond to the images when contacting the correct picture, a partial rewarding schedule was applied during the screen test (first and every third correct contact with the target image rewarded). Horses were always rewarded if they contacted the correct object on real object trials. Following the last stimulus trial (trial 15), all horses received one last target trial (single object, not included in results) to ensure that all animals ended the testing with a positive experience. Horse behaviour was recorded throughout with three GoPro cameras (Hero 3+), and number of correct responses later extracted from the videos. A second naïve coder analysed 30% of the screen test videos, which were selected at random (using Excel random number generator and choosing the first 8 videos after sorting in ascending order). Inter-observer reliability (Cohen's kappa) for coding the response behaviours was very high (0.94).

Fig.1 Experimental design A) 2-step objects discrimination training (ODT). Horses first learnt to contact a single rewarded object (target) with their muzzle to receive food. A second (unrewarded) object was subsequently added and horses trained to discriminate between both until it touched the correct object on ≥8 trials/10 over 3 consecutive 10-trials blocks. B) 3-step object recognition test (ORT). A pre-screen test was first conducted in the horse's stall. When ≥8 correct responses were performed, the horse was moved to the test arena (illustrated as rectangle with dashed lines) and re-tested in a pre-screen test to ensure it performed reliably in the new environment. When ≥8 correct responses were performed, the horse was immediately tested with images on the screen (indicated by rectangle with solid black lines). During the screen test, the horse was presented with the real objects on five trials interspersed between the 10 image trials to test whether it was still motivated to touch the objects, even if the images were not touched

2.4. Welfare assessment

Previous studies have suggested that welfare status can cause great individual variation in cognitive performance (reviewed in Hausberger et al. 2019). We therefore tested the effect of welfare condition, i.e. the level of provided environmental resources (*e.g.* stall space, pasture access), social factors (*e.g.* ability and stability of social contact) and animal-based measurements (including health indicators, workload, abnormal behaviour), putatively contributing to good horse welfare on learning ability and test performance.

The welfare assessment protocol was developed as part of another study (Kappel et al. *in prep*). Details to the protocol are provided in the Supplementary Information (Table 2). Briefly, for each factor, non-weighted numerical scores were given (0-1 indicating absence/presence of resource) and all scores combined to calculate an overall welfare score (maximum score was 20 with higher scores reflecting better welfare conditions).

2.5. Statistical analysis

Horses' responses to the objects/images were extracted from footage and coded as "correct" if the horses touched the rewarded object/image, and "wrong" if the unrewarded object/image or if neither object/image was touched. Hence, responses in trials where horses did not react to the stimuli were counted as "wrong" outcome. Furthermore, the location (left/right) of the target image was recorded to assess side effects.

Data were analysed in R (R Core team 2021). Age and the welfare scores of horses between the yards were compared using Wilcoxon rank sum tests. The number of trial blocks needed to reach learning criterion in ODT was assessed as a measure of learning ability and followed a normal distribution (Shapiro- Wilk's test, p=0.09). Thus the effect of fixed factors (*i.e.* yard, target) and covariates (*i.e.* age, welfare score) on learning ability were assessed by fitting generalised linear models (glm() function with Gaussian distribution in lme4 package, Bates et al.). Predictor covariance was check with the vif() function from the car package (Fox and Weisberg 2019), which indicate that age co-varied with the other fixed factors (vif=7.08). The effect of age on learning ability was therefore separately analysed using Pearson correlation test. Sex was not used as fixed factor given the unbalanced number of females (n=10) and males (n=17) in the final sample of horses.

Indication of recognition ability at group level was assessed by measuring whether the number of horses responding correctly and incorrectly on trial 1 of the screen test was significantly different from random using a Chi-square test. To test if the proportion of correct responses performed at group level in each of the ORT tests (*i.e.* pre-test, pre-screen test and screen test) was better than chance, one-sample Wilcoxon tests were used. Whether proportions of correct responses differed between trials following real object trial and trials following image trial was tested with a Chi-square test. Likewise, we tested the effect of reward delivery (i.e. received or withheld upon correct image contact) on subsequent trial performance using a Chi-square test.

Individual performance (correct/wrong response) during the 10 image trials was modelled using generalised linear mixed models (GLMMs; glmer() function in lme4 package, binomial family) with target type (kong^{TM}/ring), target side (left/right), and trial order (after object/not after object) as categorical fixed factors, age and welfare score as covariates, and horse ID as random factor. P-values were exacted via the anova() function from the car package and reported as significant for $p \leq 0.05$ and as trends for $p < 0.1$.

3. Results

3.1. Learning ability during object discrimination training

In total, 27 horses (16 out of 17 at yard A, 11 out of 16 at yard B) learnt to discriminate between the two objects. Overall, horses needed 11 trial blocks (median, Q1-Q3=7-15) to reach learning criterion. Learning ability was predicted by target and yard, with horses from yard B (vs yard A) and those trained with the ring (vs with the kong^{TM}) needing more trials, but by not welfare level (see Table 1 for model estimates). Pearson correlation test indicated a significant positive correlation between learning ability and age (t_{25}=4.09, r^2=0.63, p=0.0003). Horses from yard A were significantly younger (mean±SD, 10.6±2.51; W=3950, p<0.0001) and had significantly lower welfare scores (14.1±1.30, W=4400, p<0.0001) than horses from yard B (age: 14.8±5.7, welfare score: 15.5±1.73).

Table 1 Estimated regression parameters from the GLM model. Learning ability (dependent variable) predicted by welfare score, yard and target type with information to the comparator category in square brackets. P-values ≤0.05 are highlighted in bold

Predictors	Estimates	Confidence Interval	P-value
(Intercept)	1.08	0.98 – 1.20	0.146
welfare score	1	1.00 – 1.01	0.326
yard [B]	0.97	0.94 – 0.99	**0.008**
target [ring]	0.96	0.94 – 0.99	**0.005**
Observations	27		
R^2	0.454		

3.2. Objects Recognition Test

3.2.1. Image recognition at first presentation

When the horses were first presented with the images, 92.6 % of the horses (25/27) spontaneously reacted to the images as trained, *i.e.* by contacting one of the two objects' images with their nose. However, the number of horses responding correctly by touching the target image (n=14) was not significantly different from the number of horses responding incorrectly (combining the 11 horses that contacted the image of the unrewarded object, and the 2 horses that did not contact the screen at all; X^2_1=0.03, p=0.8).

3.2.2. Performance during the different stages of the ORT

Fig.2 shows the proportion of correct responses during the pre-test (PT) and pre-screen test (PST) leading up to the screen test. Since all horses needed to perform at least 8 out of 10 responses in the PT to move on to the PST, and in the PST to be tested with the images on screen (which all horses did, although some animals were re-tested in PST, see Table 3 in Supplementary Information), the effect of fixed factors (*i.e.* target, age, welfare score) on individual performance in the PT and PST tests was not further analysed. At group level, horses performed significantly better than chance (50%, V=36585, p<0.0001, see Fig.2) in PT, PST (as required) and on object trials, but significantly below this threshold during image trials (V=7340, p<0.0001).

Considering individual performance over the 10 image trials, one horse performed above chance level by selecting the correct target images 9 times (p=0.021). Other horses (N=3) always contacted the correct image when making contact with the screen, but failed to touch the images on other trials (two horses did not touch the screen on four trials, one on two trials), and therefore were not considered to perform better than chance (6/10 and 8/10 correct, both p>0.1). An overview of individuals' performance when omitting trials where horses did not make any image contacts is provided in the Supplementary Information Table 4.

Fig. 2 Proportion of correct responses during each step of the object recognition test (ORT). The results of the screen test are shown separately for the 10 images trials ('screen test (images)') and 5 real objects trials interspersed between image trials ('screen test (object)'). Dashed line indicates 50 % correct (chance level performance) against which group level performance was tested (one-sample Wilcoxon test, *** p<0.0001 (note that performance above chance level during PT and PST was required for the horses to move the screen test). Horses significantly performed below the 50 % threshold during screen test with images. Lines across boxplots show individual performances throughout the stages of the ORT. One horse touched the correct images significantly above chance level during screen test (images); data for this individual is indicated as bold line

3.2.3. Factors influencing response to the images

Horses' response to the images (*i.e.* correct/wrong) was predicted by the type of preceding trial (p<0.001, model estimates shown in Table 2). Horses were more likely to respond correctly in trials following real object trials than in trials following images trials (X^2_1=8.45, p=0.004), although the proportion of horses touching the correct image was only 51.8% (Fig.3). Overall, horses did not make any image contacts on 144 trials (53.3%), whereas horses always approached the real-objects.

Whether horses received a reward upon correct image contact or reward was (unexpectedly) withheld during a preceding image trial had no significant effect on horses' performance ($X^2_2=0.268$, p=0.874). However, images on the right side more like to result in correct responses than when the target was shown on the left ($X^2_1=3.85$, p=0.05, model estimates in Table 2).

Fig. 3 Proportion of horses out of the 27 horses responding correctly or incorrectly depending on whether the preceding trial was refreshed with objects (yes) or not (no). More horses performed correctly than incorrectly when the preceding trial was refreshed with objects (p=0.004)

Table 2 Model estimates of GLMM with response as binary dependent variable (correct/wrong) and predictors with comparator information in square brackets. P-values ≤0.05 are highlighted in bold

Predictors	Odds Ratios	Confidence Interval	P-Value
(Intercept)	0.24	0.00 – 47.44	0.6
Yard [B]	2.12	0.60 – 7.51	0.246
Age	1.09	0.94 – 1.27	0.254
Welfare score	1.06	0.76 – 1.48	0.716
Side [R]	1.79	1.00 – 3.19	**0.05**
Target [ring]	0.54	0.17 – 1.74	0.304
AfterObject [yes]	0.33	0.18 – 0.61	**<0.001**
Learning ability	1.01	0.87 – 1.19	0.869
Marginal R2 / Conditional R2	0.174 / 0.347		

4. Discussion

This study investigated if horses can recognise real-life objects from on-screen images. The majority of horses initially reacted to images with the conditioned response (*i.e.* touching the target with their muzzle for food), but the number of horses touching the correct image was not significantly different from the number of horses contacting the wrong image. Therefore, performance at group level did not suggest that the horses recognised the real objects from their 2D representations shown on-screen. However, we found that more correct responses being performed on image trials following real object trials, suggesting that horses' reactions to the images was not completely random. In fact, one horse selected the correct images at a level significantly above chance when tested repeatedly over 10 images trials, suggesting that this individual recognised the images either as the real object (confusion mode) or as a representation of it (equivalence mode; Fagot et al. 2010).

Previous studies have reported that horses are able to recognise other individuals from photographs (Smith et al. 2016; Wathan et al. 2016; Proops et al. 2018; Lansade et al. 2020a, b). As presented in the introduction, the validity of existing evidence might be hampered by experimental limitations (see Amici (2019) for discussion of Proops et al. (2018). Moreover, discrimination ability is not automatic proof of recognition (Aust and Huber 2006), and alternative mechanisms such as learning, categorisation (*i.e.* of biologically relevant objects such as food), or habituation might also influence animals' responses to repeated presentation with images (reviewed in Bovet and Vauclair 2000). Here we tested if horses would spontaneously respond to on-screen images with the same learnt response that they were trained to make to real objects, using a relative low number of test trials and partial reward delivery to avoid learning. In contrast to previous reports, our horses failed to recognise the objects from images, except for one individual. Several aspects need to be considered to put our findings in context with previous findings.

When exposed to the images for the first time, all but two horses spontaneously responded to the images with the conditioned response, suggesting the horses made some association between images and objects since the stimuli provoked the learnt behaviour. We trained the horses to express their choice by contacting the target with their muzzle, because this conditioned behaviour is commonly used in horses tested in two-choice discrimination tests (*e.g.* Flannery 1997; Hanggi 2001, 2003; Lansade et al. 2020a, b). In retrospect, we question the suitability of this behaviour as conditioned response. Horses naturally use their nose to explore unfamiliar items to gather olfactory/tactile information whilst inspecting novel objects (De Boyer Des Roches et al. 2008). Therefore, the horses might have contacted the images to explore the items rather than performing a conditioned behaviour. This might explain why we found no significant preference for either image at first presentation (trial 1). Utilising stimulus specific adaptive responses as done in studies in other species (*e.g.* grasping behaviour in marmosets (Oh et al. 2019), eating attempts of banana images in gorillas (Parron et al. 2008), or shaping behaviours distinctively different from normal horse behaviour (*e.g.* level pressing; Dougherty and Lewis 1991) could avoid this problem of ambiguity.

Intriguingly, horses were nevertheless more likely to make correct responses to the images following real object trials than following image trials. Maybe responding to the real objects before seeing the images somehow facilitated horses' ability to transfer between the stimuli, despite perceptual differences (*e.g.* lack of depth cues), for instance by matching them based on relational sameness (*e.g.* shape). In fact, Flannery (1997) observed that horses have the capacity to learn higher-order discriminations based on relation between stimuli, such as geometric shapes. It could be that horses initially confused objects and images (*i.e.* seeing both as the same), but once they made physical contact with the images, the mismatch in sensory feedback (*e.g.* olfactory/tactile feedback) between the familiar object and images resulted in independent processing of both as completely different items. Moreover, cross-modal differences (*i.e.* looks like target but does not smell/feel like target), might have stopped the horses from touching the images. Horses use cross-modal (visual/olfactory and auditory information) sensory input to recognise individuals (*e.g.* horses (Proops et al. 2009); humans (Lampe and Andre 2012; Proops et al. 2013)), but whether this is also true for identifying (familiar) objects has not been tested yet.

In addition, other experimental limitations might have influenced our findings. Work by one other group used digital stimuli (computer screens (Lansade et al. 2020a, b) and projections (Trösch et al. 2019, 2020)) which is why we anticipated that this type of visual information would be suitable for the purpose of our study. However, image quality and differences in colour perception of the images resulting from the use of the LCD projector (images generated from a light signal comprised of red, blue and green components but horses cannot perceive red/green colours) may have contributed to sensory image impressions different in horses to those generated by the real object, and to what humans see in digital images. Besides, the equine eye is adapted to dim light conditions and scattered light (*e.g.* from a bright light source such as a projector) can lead to loss of resolution (Hebel 1976). One may wonder whether the close distance to the screen might have hindered our horses' ability to clearly see the items in front of them given the blind spot directly in front of their forehead and limited visual acuity in close proximity (Hebel 1976; Timney and Macuda 2001, reviewed in (Rørvang et al. 2020). Our setup seems appropriate since others reported that horses successfully learn to discriminate between symbols of difference shapes and sizes, and photographs, when standing directly (≤50 cm) in front of a screen and contacting the stimuli with their muzzle (Gabor and Gerken 2012; Tomonaga et al. 2015). Varying the blinded handler position (always positioned on the left-hand side for practical reasons) should be considered for future work, since we found that targets presented on the right side were more likely to result in correct responses. The spatial relationship between cue, reward and response influences discrimination learning (Miller and Murphy 1964; Hothersall et al. 2010), which might explain why target location tended to affect performance.

Maybe our results do not support previous reports of image recognition in horses because of the type of stimuli we used. From an adaptive perspective, processing visual cues of biological relevance is highly important, and images representing biologically relevant stimuli (*e.g.* prey, conspecifics, predators) are instrumental in studies of animal picture recognition where animals' spontaneous (initial) response to pictorial cues is tested (Bovet and Vauclair 2000). For instance, (Kendrick et al. 1996) observed that sheep were much faster in learning to discriminate between images of conspecifics (familiar or unfamiliar) than between geometrical shape discrimination cues, possibly because sheep seem to cognitively process information associated with social familiarity (*i.e.* facial features of conspecifics) more efficiently than non-social cues. It seems probable that specialised sensory processing of social cues is also relevant to horses, since they show a range of postural and facial expressions for social communication (Waring 2003; Wathan et al. 2015), and understand visual cues from humans (Proops and McComb 2010). It seems therefore possible that equine studies using images of conspecifics (Wathan et al. 2016) or humans (Smith et al. 2016; Proops et al. 2018; Lansade et al. 2020b, a) tap into different sensory processing levels than when objects images are used. We chose real-life objects instead of images of conspecifics as this allowed us to train and test horses' response more easily under controlled conditions (*i.e.* excluding variation within the test stimuli). We also excluded food cues since disentangling animals' motivation to respond to food cues when food rewards are provided during repeated testing might be difficult. Nevertheless, we expected that the horses would pay attention to the on-screen stimuli if they perceive the images as equal to the real objects given that they had learnt to associate these with food (*i.e.* a biologically relevant resource).

Digital images are increasingly applied in the study of horse cognition, but evidence that this species has the ability to recognise the content of digital images is still sparse. Hence, we investigated how horses' spontaneously respond to on-screen images of known objects and did not consider to test the reverse (i.e. whether horses' recognise real-life objects from images). We do encourage future research to study this further (considering cofounding factors discussed in the introduction regarding Hanggi et al., 2001), for instance, to understand what stimulus characteristics (e.g. colour (Hangii and Ingersoll, 2007), shape, or size (Tomonaga et al., 2015; Hanggi 2003)) drive recognition as these could be easily manipulated in digital images. Here, we only used two real-life objects distinctively differing in colour, shape and size (stimulus features horses can generally discriminate) as using more items could have introduced more variability in individuals' responses making the interpretation of findings more difficult. We therefore believe our findings that horses overall did not perform reliably enough to suggest image recognition using two objects are of significance. However, we must acknowledge that our observations may not be generalizable as the use of different objects could have led to different findings.

Only one out of 27 horses responded to the stimuli on screen above chance level suggesting that this individual might have recognised the images as objects or representations of such. Rapid learning seems unlikely given the experimental precautions we undertook. For example, we used partial reinforcement in the screen test to reduce the possibility that horses would respond to image-related cues, *i.e.* exhibit the muzzle contact as new behaviour specific to the images rather than touching them because they were recognise as a replacement of the objects. Indeed, horses were not more likely to respond correctly to the images following images trials a reward was delivered upon correct response than when reward was (unexpectedly) withheld. Likewise, horses were not more likely to respond incorrectly following unrewarded trials. Still, learning cannot be completely excluded as partial reinforcement reduces, but does not exclude, acquisition of a conditioned response compared to continuous reinforcement (Gottlieb 2005; Amselme 2014). On the other hand, the performance of this particular horse might reflect a statistical Type I error (Pollard & Richardson 1987). Our study design does not allow us to draw conclusions as to *how* this horse recognised the images (*i.e.* whether images and objects were seen as the same item (*i.e.* confused), the images seen as functional representations (equivalent) of the target or both processed independently). Nevertheless, this finding is interesting as it highlights the importance of considering individual variation in cognitive tests. In correspondence with other findings showing that older horses learn more slowly in a social learning task (Krueger et al. 2014), we found that older horses needed more trial blocks to learn the discrimination task, but we found no association between welfare level, learning ability and test performance. Further study could investigate further inter-individual differences such as variations in personality or in perceptual abilities on performance.

5. Conclusion

Only one of 27 horses responded to the images suggesting it might have recognised the images as objects or representations of such, while all other horses apparently failed to do so. As a species, horses may possess the basic capability to perceive the content of artificial visual stimuli such as digital image, but our findings indicate that in horses unfamiliar with two-dimensional representations image recognition might not be an ability that can be generalised across horses and testing situations. Instead, further research is warranted in order to understand how horses perceive (at sensory level) and interpret (at cognitive level) images for the human eye, especially if they are to be utilised as representations of real-life objects, as well as inter-individual variations in such abilities. Until then, we do not know if humans and horses see *eye to eye* when viewing this type of artificial stimuli.

Acknowledgements:

We would like to thank the riding schools for the use of their animals and facilities. This project was founded by the University of Plymouth.

References:

Amici F (2019) Memories of emotional expressions in horses. Learn Behav 47:191–192. https://doi.org/10.3758/s13420-018-0363-9

Anselme P (2015) Incentive salience attribution under reward uncertainty: A Pavlovian model. Behavioural Processes 111:6–18. https://doi.org/10.1016/j.beproc.2014.10.016

Aust U, Huber L (2006) Picture-object recognition in pigeons: Evidence of representational insight in a visual categorization task using a complementary information procedure. Journal of Experimental Psychology: Animal Behavior Processes 32:190–195. https://doi.org/10.1037/0097-7403.32.2.190

Bates D, Maechler M, Bolker B, Walker S Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software 67:. https://doi.org/10.18637/jss.v067.i01

Bovet D, Vauclair J (2000) Picture recognition in animals and humans. Behavioural Brain Research 109:143–165. https://doi.org/10.1016/S0166-4328(00)00146-7

Cabe PA (1976) Transfer of discrimination from solid objects to pictures by pigeons: A test of theoretical models of pictorial perception. Perception & Psychophysics 19:545–550. https://doi.org/10.3758/BF03211224

Carere C, Locurto C (2011) Interaction between animal personality and animal cognition. Current Zoology 57:491–498. https://doi.org/10.1093/czoolo/57.4.491

Christensen JW, Ahrendt LP, Lintrup R, et al (2012) Does learning performance in horses relate to fearfulness, baseline stress hormone, and social rank? Applied Animal Behaviour Science 140:44–52. https://doi.org/10.1016/j.applanim.2012.05.003

De Boyer Des Roches A, Richard-Yris M-A, Henry S, et al (2008) Laterality and emotions: Visual laterality in the domestic horse (Equus caballus) differs with objects' emotional value. Physiology & Behavior 94:487–490. https://doi.org/10.1016/j.physbeh.2008.03.002

D'Eath RB (1998) Can video images imitate real stimuli in animal behaviour experiments? Biol Rev 73:267–292. https://doi.org/10.1017/S0006323198005179

Dougherty DM, Lewis P (1991) Stimulus Generalization, Discrimination Learning, and Peak Shift in Horses. Journal of the Experimental Analysis of Behavior 56:97–104. https://doi.org/10.1901/jeab.1991.56-97

Dougherty LR, Guillette LM (2018) Linking personality and cognition: a meta-analysis. Phil Trans R Soc B 373:20170282. https://doi.org/10.1098/rstb.2017.0282

Duke-Elder S System of ophthalmology: vol. I. The eye in evolution. Henry Kimpton, London

Fagot J (2000) Picture Perception in Animals. Psychology Press, East Sussex

Fagot J, Parron C (2010) Picture Perception in Birds: Perspective from Primatologists. CCBR 5:132–135. https://doi.org/10.3819/ccbr.2010.50007

Fagot J, Thompson RKR, Parron C (2010) How to read a picture: Lessons from nonhuman primates. Proceedings of the National Academy of Sciences 107:519–520. https://doi.org/10.1073/pnas.0913577107

Flannery B (1997) Relational discrimination learning in horses. Applied Animal Behaviour Science 54:267–280. https://doi.org/10.1016/S0168-1591(97)00006-3

Fox J, Weisberg S (2019) An {R} Companion to Applied Regression. Version Third Edition. Sage, Thousand Oaks CA. URL https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gabor V, Gerken M (2012) Cognitive testing in horses using a computer based apparatus. Applied Animal Behaviour Science 139:242–250. https://doi.org/10.1016/j.applanim.2012.04.010

Gottlieb DA (2004) Acquisition with partial and continuous reinforcement in pigeon autoshaping. Animal Learning & Behavior 32:321–334. https://doi.org/10.3758/BF03196031

Grzimek B (1943) Begrussung zweier Pferde: Das Erkennen von Phantomen und Bildern. 465–480

Hanggi E (2001) Can horses recognize pictures? In: Proceedings of the third international conference of cognitive science, Beijing, China, pp 52–56

Hanggi EB (1999) Categorization learning in horses (Equus caballus). Journal of Comparative Psychology 113:243–252. https://doi.org/10.1037/0735-7036.113.3.243

Hanggi EB (2003) Discrimination learning based on relative size concepts in horses (Equus caballus). Applied Animal Behaviour Science 83:201–213. https://doi.org/10.1016/S0168-1591(03)00136-9

Hanggi EB, Ingersoll JF (2009a) Stimulus discrimination by horses under scotopic conditions. Behavioural Processes 82:45–50. https://doi.org/10.1016/j.beproc.2009.04.009

Hanggi EB, Ingersoll JF (2009b) Long-term memory for categories and concepts in horses (Equus caballus). Anim Cogn 12:451–462. https://doi.org/10.1007/s10071-008-0205-9

Hanggi EB, Ingersoll JF, Waggoner TL (2007) Color vision in horses (Equus caballus): Deficiencies identified using a pseudoisochromatic plate test. Journal of Comparative Psychology 121:65–72. https://doi.org/10.1037/0735-7036.121.1.65

Hausberger M, Stomp M, Sankey C, et al (2019) Mutual interactions between cognition and welfare: The horse as an animal model. Neuroscience & Biobehavioral Reviews 107:540–559. https://doi.org/10.1016/j.neubiorev.2019.08.022

Hebel R (1976) Distribution of retinal ganglion cells in five mammalian species (pig, sheep, ox, horse, dog). Anat Embryol 45–51

Hothersall B, Gale EV, Harris P, Nicol CJ (2010) Cue use by foals (Equus caballus) in a discrimination learning task. Anim Cogn 13:63–74. https://doi.org/10.1007/s10071-009-0245-9

Hughes A (1977) The Topography of Vision in Mammals of Contrasting Life Style: Comparative Optics and Retinal Organisation. In: Crescitelli F (ed) The Visual System in Vertebrates. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 613–756

Jacobs GH, Deegan JF, Neitz J (1998) Photopigment basis for dichromatic color vision in cows, goats, and sheep. Vis Neurosci 15:. https://doi.org/10.1017/S0952523898153154

Kendrick KM, Atkins K, Hinton MR, et al (1996) Are faces special for sheep? Evidence from facial and object discrimination learning tests showing effects of inversion and social familiarity. Behavioural Processes 38:19–35. https://doi.org/10.1016/0376-6357(96)00006-X

Krueger K, Farmer K, Heinze J (2014) The effects of age, rank and neophobia on social learning in horses. Anim Cogn 17:645–655. https://doi.org/10.1007/s10071-013-0696-x

Lampe JF, Andre J (2012) Cross-modal recognition of human individuals in domestic horses (Equus caballus). Anim Cogn 15:623–630. https://doi.org/10.1007/s10071-012-0490-1

Lansade L, Colson V, Parias C, et al (2020a) Female horses spontaneously identify a photograph of their keeper, last seen six months previously. Sci Rep 10:6302. https://doi.org/10.1038/s41598-020-62940-w

Lansade L, Colson V, Parias C, et al (2020b) Human Face Recognition in Horses: Data in Favor of a Holistic Process. Front Psychol 11:575808. https://doi.org/10.3389/fpsyg.2020.575808

Miller RE, Murphy JV (1964) Influence of the spatial relationships between the cue, reward, and response in discrimination learning. Journal of Experimental Psychology 67:120–123. https://doi.org/10.1037/h0040113

Murphy J, Waldmann T, Arkins S (2004) Sex differences in equine learning skills and visuo-spatial ability. Applied Animal Behaviour Science 87:119–130. https://doi.org/10.1016/j.applanim.2003.12.002

Nakamura, K., Takimoto-Inose, A., & Hasegawa, T. (2018). Cross-modal perception of human emotion in domestic horses (Equus caballus). Scientific Reports, 8(1), 8660. https://doi.org/10.1038/s41598-018-26892-6

Oh J, Šlipogor V, Fitch WT (2019) Artificial visual stimuli for animal experiments: An experimental evaluation in a prey capture context with common marmosets (Callithrix jacchus). Journal of Comparative Psychology 133:72–80. https://doi.org/10.1037/com0000129

Parron C, Call J, Fagot J (2008) Behavioural responses to photographs by pictorially naïve baboons (Papio anubis), gorillas (Gorilla gorilla) and chimpanzees (Pan troglodytes). Behavioural Processes 78:351–357. https://doi.org/10.1016/j.beproc.2008.01.019

Pfungst O, Rahn CL (1911) Clever Hans: the horse of Mr. Von Osten. A contribution to experimental animal and human psychology. Holt, Rineheart, and Winston, New York

Pollard, P., & Richardson, J. T. (1987). On the probability of making Type I errors. Psychological Bulletin, 102(1), 159–163. https://doi.org/10.1037/0033-2909.102.1.159

Proops L, Grounds K, Smith AV, McComb K (2018) Animals Remember Previous Facial Expressions that Specific Humans Have Exhibited. Current Biology 28:1428-1432.e4. https://doi.org/10.1016/j.cub.2018.03.035

Proops L, McComb K (2010) Attributing attention: the use of human-given cues by domestic horses (Equus caballus). Anim Cogn 13:197–205. https://doi.org/10.1007/s10071-009-0257-5

Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). PNAS 106:947–951. https://doi.org/10.1073/pnas.0809127105

Proops L, Rayner J, Taylor AM, McComb K (2013) The Responses of Young Domestic Horses to Human-Given Cues. PLoS One 8:e67000. https://doi.org/10.1371/journal.pone.0067000

R Core team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rørvang MV, Nielsen BL, McLean AN (2020) Sensory Abilities of Horses and Their Importance for Equitation Science. Front Vet Sci 7:633. https://doi.org/10.3389/fvets.2020.00633

Rowe C, Healy SD (2014) Measuring variation in cognition. Behavioral Ecology 25:1287–1292. https://doi.org/10.1093/beheco/aru090

Saslow CA (2002) Understanding the perceptual world of horses. Applied Animal Behaviour Science 78:209–224. https://doi.org/10.1016/S0168-1591(02)00092-8

Smith AV, Proops L, Grounds K, et al (2016) Functionally relevant responses to human facial expressions of emotion in the domestic horse (*Equus caballus*). Biol Lett 12:20150907. https://doi.org/10.1098/rsbl.2015.0907

Timney B, Macuda T (2001) Vision and hearing in horses. Journal of the American Veterinary Medical Association 218:1567–1574. https://doi.org/10.2460/javma.2001.218.1567

Tomonaga M, Kumazaki K, Camus F, et al (2015) A horse's eye view: size and shape discrimination compared with other mammals. Biol Lett 11:20150701. https://doi.org/10.1098/rsbl.2015.0701

Trösch, Cuzol, Parias, et al (2019) Horses Categorize Human Emotions Cross-Modally Based on Facial Expression and Non-Verbal Vocalizations. Animals 9:862. https://doi.org/10.3390/ani9110862

Trösch M, Pellon S, Cuzol F, et al (2020) Horses feel emotions when they watch positive and negative horse–human interactions in a video and transpose what they saw to real life. Anim Cogn 23:643–653. https://doi.org/10.1007/s10071-020-01369-0

Valenchon M, Lévy F, Prunier A, et al (2013) Stress Modulates Instrumental Learning Performances in Horses (Equus caballus) in Interaction with Temperament. PLOS ONE 8:e62324. https://doi.org/10.1371/journal.pone.0062324

Vandenheede M, Bouissou MF (1994) Fear reactions of ewes to photographic images. Behavioural Processes 32:17–28. https://doi.org/10.1016/0376-6357(94)90024-8

von Borell E, Langbein J, Després G, et al (2007) Heart rate variability as a measure of autonomic regulation of cardiac activity for assessing stress and welfare in farm animals — A review. Physiology & Behavior 92:293–316. https://doi.org/10.1016/j.physbeh.2007.01.007

Waring GH (2003) Horse Behaviour, 2nd edn. Noyes Publications, New York

Wathan J, Burrows AM, Waller BM, McComb K (2015) EquiFACS: The Equine Facial Action Coding System. PLoS ONE 10:e0131738. https://doi.org/10.1371/journal.pone.0131738

Wathan J, Proops L, Grounds K, McComb K (2016) Horses discriminate between facial expressions of conspecifics. Sci Rep 6:38322. https://doi.org/10.1038/srep38322

Weisman R, Spetch M (2010) Determining When Birds Perceive Correspondence Between Pictures and Objects: A Critique. CCBR 5:117–131. https://doi.org/10.3819/ccbr.2010.50006