Animal Cognition (2019) 22:17–33
https://doi.org/10.1007/s10071-018-1217-8

ORIGINAL PAPER

Stereotypic horses (*Equus caballus*) are not cognitively impaired

Sabrina Briefer Freymond[1] · **Alice Ruet**[1] · **Maurine Grivaz**[1] · **Camille Fuentes**[1] · **Klaus Zuberbühler**[2,4] · **Iris Bachmann**[1] · **Elodie F. Briefer**[3]

Received: 4 July 2018 / Revised: 28 September 2018 / Accepted: 8 October 2018 / Published online: 17 October 2018

Abstract

Stereotypies in animals are thought to arise from an interaction between genetic predisposition and sub-optimal housing conditions. In domestic horses, a well-studied stereotypy is crib-biting, an abnormal behaviour that appears to help individuals to cope with stressful situations. One prominent hypothesis states that animals affected by stereotypies are cognitively less flexible compared to healthy controls, due to sensitization of a specific brain area, the basal ganglia. The aim of this study was to test this hypothesis in crib-biting and healthy controls, using a cognitive task, reversal learning, which has been used as a diagnostic for basal ganglia dysfunction. The procedure consisted of exposing subjects to four learning tasks; first and second acquisition, and their reversals. For each task, we measured the number of trials to reach criterion and heart rate and heart-rate variability. Importantly, we did not try to prevent crib-biters from executing their stereotypic behaviour. We found that the first reversal learning task required the largest number of trials, confirming its challenging nature. Interestingly, the second reversal learning task required significantly fewer trials to reach criterion, suggesting generalisation learning. However, we did not find any performance differences across groups; both stereotypic and control animals required a similar numbers of trials and did not differ in their physiological responses. Our results thus challenge the widely held belief that crib-biting horses, and stereotypic animals more generally, are cognitively impaired. We conclude that cognitive underperformance may occur in stereotypic horses if they are prevented from crib-biting to cope with experienced stress.

Keywords Crib-biting · Basal ganglia · Learning capacity

Introduction

Stereotypies exist in humans and non-human animals and have been defined as repetitive, relatively invariant patterns of behaviour with no apparent goal or function (Mason and Latham 2004). In animals, they seem to be artefacts of a captive environment involving restricted and sub-optimal housing conditions (McGreevy et al. 1995). They exist in various forms, the most common forms being oral and locomotor stereotypies (Mason and Rushen 2006). One classic ethological model of motivation suggests that restrictive environments can prevent the animal from reaching the "consummatory phase" of a behaviour (e.g., feeding) (Hughes and Duncan 1988). As a consequence, a number of appetitive behaviours (e.g., locomotor action to access food) are being attempted in an effort to reach unobtainable goal states (e.g., elevated blood glucose). When the goal is available, its consummation results in a negative feedback that decreases motivation. In contrast, in cases where the goal is unachievable, the lack of consummation and resulting absence of negative feedback increases the level of motivation to perform appetitive behaviours. If such frustration-inducing situations occur repeatedly, stereotypies can develop (Mason and Rushen 2006; McBride and Parker 2015). Along similar lines of reasoning, it has been suggested that privation (as experienced in captivity) increases the desire for rewards, which in turn enhances appetitive behaviours such as locomotor or feeding behaviours (Spruijt et al. 2001). Excessive use of appetitive behaviours, can, in turn, lead to a loss of

---

✉ Sabrina Briefer Freymond
sabrina.briefer@agroscope.admin.ch

[1] Agroscope, Swiss National Stud Farm SNSTF, Les Longs Prés, 1580 Avenches, Switzerland

[2] Institute of Biology, University of Neuchatel, rue Emile-Argand 11, 2000 Neuchâtel, Switzerland

[3] Institute of Agricultural Sciences, ETH Zürich, Universitätsstrasse 2, 8092 Zurich, Switzerland

[4] School of Psychology and Neuroscience, University of St Andrews, St Andrews, Scotland, UK

regulatory control, which means that these behaviour will become environmentally irreversible (inflexible) and can develop into stereotypies (Toates 2004).

Other studies have focussed on the impact of chronic stress on the brain, and in particular, on parts of the basal ganglia (the mesoaccumbens dopaminergic system) (Spruijt et al. 2001). The reasoning behind this research is that stressful negative experiences lead to an increased tendency to seek rewards via a neurobiological process of sensitisation (van der Harst et al. 2003), a possible common principle underlying stereotypies in evolutionarily distant species such as mice and human (Cabib et al. 1998). For example, some studies have shown an association between stressors, stereotypy development and dopamine receptor function in the basal ganglia in rodents (Cabib et al. 1998) and crib-biting horses, *Equus caballus* (McBride and Hemmings 2005), suggesting that the basal ganglia plays a general role in the development of stereotypic behaviour (McBride et al. 2017). However, the exact causal relationship between basal ganglia alterations and the emergence of stereotypic behaviour is currently unknown (McBride and Hemmings 2005).

Each part of the basal ganglia plays a specific role during instrumental learning (Dickinson 1985). The first stage ("acquisition") is mediated by the ventral and dorsomedial striatum of the basal ganglia (McBride et al. 2017; Parker et al. 2008; Yin and Knowlton 2006), the second stage ("action-outcome") by the dorsomedial striatum (McBride et al. 2017; Parker et al. 2008; Yin and Knowlton 2006), and the final stage ("habit formation") by a shift in activation from the dorsomedial towards the dorsolateral striatum of the basal ganglia. To summarize, the ventral striatum and dorsomedial striatum seem to be important in the acquisition and execution of goal-directed actions, that is, in establishing the link between stimulus response and outcome (McBride et al. 2017; Yin et al. 2008). By contrast, the dorsolateral striatum seems to control subsequent habit formation, independently of the response outcome (Yin and Knowlton 2006). Therefore, due to its primary functional role, dysfunctions in the basal ganglia might lead to behavioural abnormalities linked to impaired action selection and also to impairments in controlling instrumentally learned behaviour (McBride and Hemmings 2005; Vickery and Mason 2005).

Although there are strong reasons to assume that an impaired basal ganglia function is related to the development of stereotypic behaviour, research is challenging for financial, logistical, and ethical reasons, which has led to the development and use of indirect and non-invasive methods (McBride et al. 2017). For instance, stereotypy levels (frequency of stereotypy performance) have been shown to constitute strong predictors of the latency to extinguish conditioned responses or of the tendency to inappropriately repeat responses, both of which constitute indirect measures of basal ganglia dysfunction in several species such as bears, *Ursus thibetanus* and *Helarctos malayanus* (Vickery and Mason 2005), Orange-winged amazon parrots, *Amazona amazonica* (Garner et al. 2003) and bank voles, *Clethrionomys glareolus* (Garner and Mason 2002). In humans, poor abilities to suppress learnt behaviour (perseveration) have been shown in autistic patients who are prone to stereotypic behaviour (Boyd et al. 2009; Lopez et al. 2005).

Domesticated horses are subject to management practices that make them prone to develop stereotypic behaviours. Understanding the nature of stereotypies and their impact on learning abilities is therefore of considerable importance for horse owners. Crib-biting, an oral stereotypy, is one of the most common forms of stereotypy in horses (Luescher et al. 1991; Wickens and Heleski 2010). The performance of this behaviour varies between horses in terms of the percentage of time occupied by the stereotypic behaviour (Houpt and McDnnell 1993). Crib-biting has been linked to learning impairments in extinction paradigms (Hemmings et al. 2007; Roberts et al. 2015). In particular, stereotypic horses need more trials compared to healthy individuals before extinction of a previously learnt action occurs, and this may be linked to a basal ganglia dysfunction. In one study, crib-biting horses appeared to exhibit altered dopamine receptor sensitivity in the basal ganglia (McBride and Hemmings 2005), due to the higher number of dopamine receptors in the ventral striatum and the lower number of receptors in the dorsomedial striatum. Since the dorsomedial striatum mediates action–outcome learning, it is possible that crib-biting horses are simply unable to maintain this type of learning and show an accelerated shift from action–outcome learning to habit formation (Parker et al. 2008, 2009; Roberts et al. 2015), and a reduced ability to learn about outcomes (Schwabe and Wolf 2011). Additionally, another study including many different kinds of stereotypies in horses (locomotor and oral) showed that stereotypic horses need more time to learn an instrumental task (opening a chest by raising the lid using the nose) compared to non-stereotypic horses (Hausberger et al. 2007).

As described above, previous studies have found differences in the learning capacities of crib-biting and control horses (Hemmings et al. 2007; Parker et al. 2008, 2009; Roberts et al. 2015), and other work has shown that crib-biting horses appear to have alterations in the dopaminergic system (McBride and Hemmings 2005). However, the link between these alterations and cognitive performances has remained unclear (Roberts et al. 2017). Reversal learning paradigms are of particular relevance, as they have been used as a diagnostic tool for dopaminergic dysfunction and as general measure of cognitive flexibility in rodents, non-human primates and humans (Izquierdo et al. 2017; McBride et al. 2017). To our knowledge, reversal learning has not been investigated in crib-biting horses and has been shown to pose a challenge to this species when based on visual cues (Brubaker and Udell 2016; Hothersall et al. 2010; Martin et al. 2006; McBride et al. 2017; Sappington et al. 1997; Voith

1975), unlike reversal learning tasks based on spatial cues, which seem to be fairly easy for horses due to their ecological relevance (e.g., finding natural food sources) (Brubaker and Udell 2016; Fiske and Potter 1979; Martin et al. 2006; Voith 1975; Warren and Warren 1962). We therefore tested crib-biting and control, non-stereotypic horses in two subsequent reversal learning tasks based on visual cues. First, we predicted that crib-biting horses would need less trials than controls to perform the first and second acquisition task, because they might be more prone to habit learning than to response–outcome learning (Hemmings et al. 2007; Parker et al. 2008, 2009; Roberts et al. 2015). Second, we predicted that crib-biting horses would need more trials than controls to perform the reversal learning tasks, suggesting learning disabilities, if they suffered from dopaminergic dysfunction. By contrast, similar performances between crib-biting and control horses would suggest that the stereotypic horses are not suffering from such a dysfunction. Importantly, we did not try to prevent crib-biters from executing their stereotypic behaviour, based on our previous finding that crib-biting reduces stress (Briefer Freymond et al. 2015), to avoid any confounding influence of stress on learning (Schwabe and Wolf 2010).

Methods

Subjects and management conditions

The study was carried out on six crib-biters (“CB”) and seven control horses (“C”) ($N = 13$) of various breeds, sexes (mares, geldings and stallions), and ages (10–25 years old), housed in five different farms in Switzerland, between January and May 2016 (Table 1). Eight horses were privately owned, and five horses were owned by the Swiss National Stud Farm. All the horses had been at their respective farms for at least 1 year. To be eligible for inclusion in the study, crib-biters were required to have demonstrated crib-biting behaviour for a minimum of 1 year, as reported by their owners. All the crib-biters eventually included in the study had been crib-biting for at least 4 years. Controls were horses that had never been observed crib-biting or performing other kinds of stereotypies (e.g., weaving or box-walking). All but two animals participated in a previous study, which involved a spatial learning task (Briefer Freymond et al., in preparation; one crib-biter, one control). Each crib-biter was matched with a control horse of similar breed (except for one pair), sex, age, and housing conditions (individual or group, single box or box with paddock, and if possible in the same farm) (Table 1). One supplementary control horse was tested to lower the average age of the controls, which was originally higher than the age of the crib-biters (final mean age (years old): controls = 17.6; crib-biters = 13.5). Routine care of the study animals was provided by the farm/horse owner.

Experimental design

Experimental protocol

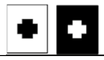
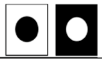
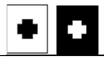
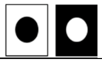
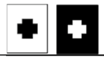
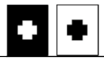
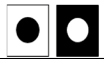
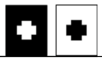
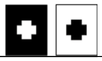
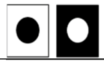
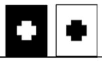
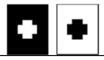
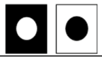
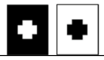
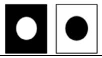
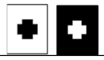
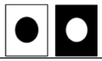
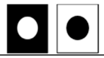
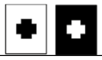
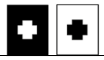
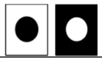
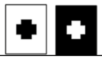
Before the start of the learning experiment, all the CB horses were filmed in their home pen, while undisturbed, during a 48-h period (excluding times when horses were ridden or in pasture), to assess their crib-biting frequency per hour (see in “Stereotypy level” below and Table 1). For the learning experiment, each horse was led individually, to a delimited (8/10 m) familiar arena, after equipping it with a heart-rate monitor in its home pen (see in “Physiological measures” below). The arena was divided into a waiting area and a test area, separated by a start rope. The learning apparatus was placed at one end of the test area (Fig. 1). The horses were filmed from the back with a video camera fixed on a pole to score their behaviour (see in “Behavioural measures” below). Two experimenters were present during the study sessions; Experimenter 1 was located in the arena and handled the subjects (see in “Discrimination procedure” below), while Experimenter 2 was located outside the arena and was entering comments on the collected physiological data (see in “Physiological measures” below) as well as preparing the learning apparatus for the next trial (see in “Discrimination procedure” below).

Apparatus

Because horses are very skilled at reading subtle unintentional human behaviour (Ringhofer and Yamamoto 2017), we built an apparatus which allowed the experimenters to remain in the back of the area (Fig. 1). Following Gabor and Gerken (2010), the apparatus consisted of a wooden box measuring 1 m (height) × 1.6 m (length) × 0.4 m (width) with two flaps (45 × 61 cm) on the front side (Fig. 1). The horse could reach the food through these flaps. To prevent the horses from using olfactory cues, Experimenter 2 always filled both bowls situated at the back side of the apparatus with food (20 g of commercial concentrate), even though only one side was rewarded each time. This was achieved by closing the unrewarded flap automatically using an electromagnet that could be activated with an infrared remote control. In addition, in case of an incorrect choice, the positively reinforced flap was immediately closed from the other side by activating the electromagnet to prevent the horse from being rewarded for an incorrect choice. A vertical piece of wood was added in the middle of the apparatus between the two flaps to better separate the two sides of the apparatus and facilitate the scoring of the horse’s choices.

The visual stimuli (see “Two-choice visual discrimination tasks” for more details) were inserted on the front side of the apparatus inside a plastic window fixed on the wooden boards. The stimuli consisted of sheets of

Table 1 Characteristics of the horses used in the experiment

Horses	Group	Breed	Sex	Birth	Acq1/Rev1	Acq2/Rev2	Colour
1	C	FM	f	1993			Y
2	C	FM	s	1996			Y
3	C	ET	g	1991			Y
4	C	FM	g	2001			Y
5	C	WB	g	2002			N
6	C	CA	f	2000			N
7	C	FM	s	2006			N
8	CB-S	FM	f	1997			Y
9	CB-M	FM	s	2002			Y
10	CB-L	FM	s	2005			Y
11	CB-M	FM	g	2004			N
12	CB-S	ET	g	2003			N
13	CB-S	HA	f	2004			N

Group (controls = "C"; crib-biters = "CB"; strong crib-biters = "S", medium crib-biters = "M", low-frequency crib-biters = "L"), Sex (female = f; gelding = g, stallion = s), Breed (Franches-Montagnes = "FM"; English thoroughbred = "ET"; warmblood = "WB"; Camargue horse= "CA"; Hispano-Arabian = "HA"), year of birth, stimuli used in the first set of learning tasks (Acquisition 1 ="Acq1″, Reversal 1 = "Rev1) and in the second set of learning tasks (Acquisition 2 = "Acq2", Reversal 2 = "Rev2"), and presence of a change in colour between Rev1 and Acq2 or not (change of colour = "Y", no change of colour = "N")

laminated paper (21 × 29.5 cm) on which either a black cross on top of a white background, a white cross on top of a black background (first set of stimuli), a black circle on top of a white background, or a white circle on top of a black background (second set of stimuli) were drawn. The same amount of sheet area was covered by the cross (13.5 × 13.5 cm) and by the circle (13 cm in diameter) (Table 1).

Acclimation and pre-training (2–6 days)

During the acclimation and pre-training phases, the horses were habituated to the experimental arena and trained to move from the starting point to the apparatus and to open the flaps in the absence of any stimuli. The horses were trained during two 10-min sessions each day, for 2–6 days. The horses were first trained to touch a target (a tennis ball fixed on a stick) with their noses and then the flaps on the apparatus, using a shaping procedure. This shaping procedure, also called "successive approximations", consisted of reinforcing behaviours directed towards the desired response (McGreevy and McLean 2010). The first step of the pre-training phase lasted until each horse was acclimated to the apparatus and touched both flaps easily with its nose. The second step consisted in shaping the horse to open the flaps, by rewarding it each time it pushed the flaps. When the horse had learnt to open the flaps with its nose, Experimenter 1 led the horse to the waiting area, and it was trained to go alone from the waiting area to open the flaps (third step). The pre-training lasted until the horses opened the flaps at least five times on both sides of the apparatus.

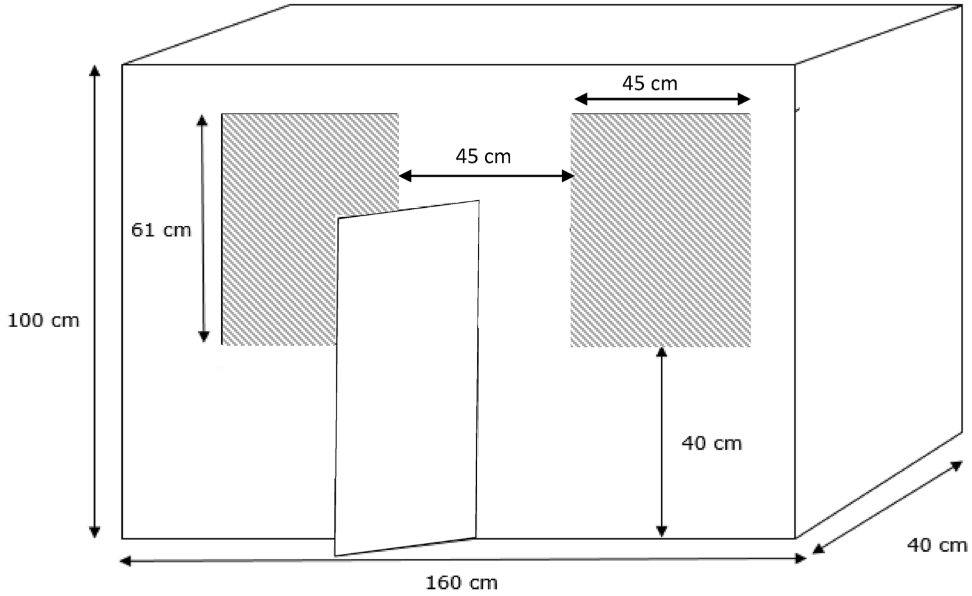


Fig. 1 Test apparatus for the visual discrimination task. The visual stimuli were inserted on the front side of the apparatus inside a plastic window (indicated with a dash line). In case of a correct choice, the horses could reach the food through the corresponding flap. A vertical piece of wood was added in the middle of the apparatus

Table 2 Number of crib-biting events performed during the tests ("CBnb") and number of trials ("Trials") needed to attain the learning criterion for the corresponding phase ("Phase"; Acq1, Rev1, Acq2, Rev2)

Horses	CBnb	Trials	Phase
8	1	144	Acq1
8	0	94	Rev1
8	0	124	Acq2
8	0	105	Rev2
9	1	109	Acq1
9	4	247	Rev1
9	0	229	Acq2
9	0	9	Rev2
10	0	14	Acq1
10	0	149	Rev1
10	0	65	Acq2
10	0	1	Rev2
11	0	74	Acq1
11	0	420	Rev1
11	0	49	Acq2
11	0	220	Rev2
12	0	111	Acq1
12	0	478	Rev1
12	0	27	Acq2
12	0	157	Rev2
13	121	53	Acq1
13	745	298	Rev1
13	380	73	Acq2
13	2634	473	Rev2

"Horses" refer to the number attributed to each horses in Table 1

Discrimination procedure

The discrimination procedure took place during the following 3–7 weeks. The horses were tested each Monday, Tuesday, Thursday and Friday. They were tested one by one with two sessions per day (15–20 min), each containing between 20 and 23 trials, with a break between the respective sessions of about 20 min, during which time another horse from the same farm was tested. Before each trial, Experimenter 2 inserted the two stimulus sheets into the plastic windows in a "pseudo-randomized" order (established a priori). This order ensured that each given stimulus was not presented for more than three consecutive trials at the same position (left or right). After inserting the stimuli sheets, the feeding bowls at the back side of the apparatus were filled and the unrewarded flap was remotely blocked. During this time, Experimenter 1 led the horse to the waiting area and released it after closing the waiting area with a rope. As soon as the setting for the next trial was ready, Experimenter 1, who was blind to the correct stimulus at the beginning of each session, opened the start rope while facing and looking away from the horse and the test area. In case of a correct choice, the horse was led back to the waiting area after reaching and eating the reward. In case of an incorrect choice, the horse was led back to the waiting area and allowed to choose again with the same arrangement of stimuli ("correction trial")(Flannery 1997). After three wrong decisions, the horse was led to the correct stimulus where it could open the flap to reach the reward (Flannery 1997). In this case, if necessary, Experimenter 1 pointed at the correct flap with the hand. Each trial was limited to a 90-s duration, during which all subjects made a choice (i.e., throughout the experiment, no trial had to be stopped before a choice was made).

Two-choice visual discrimination task

The learning procedure consisted of four different phases ("Phase"). At the start of the learning phase 1 ("Acq1"), two initial stimuli (a black versus a white cross or a black versus a white circle) were presented to the horses. The learning criterion was set at six consecutive correct responses in one learning session (i.e., probability of doing this by chance = 0.02) (McBride et al. 2016). Once the horse had reached the learning criterion for Acq1, the colour of the stimuli were reversed ("Rev1") and the next session started. Once the horses had reached the learning criterion for Rev1, they were then tested with a second set with novel stimuli (i.e., a black versus a white circle or a black versus a white cross; second acquisition phase, "Acq2"). Acq 2 was then followed by a second, final reversal ("Rev2"), after reaching the learning criterion for Acq2. Therefore, the two reversals, Rev1 and Rev2, consisted in rewarding the previously unrewarded stimulus (during Acq1 and Acq2, respectively), while Acq2 consisted in presenting circles to horses who received crosses during Acq1 and Rev1, and vice versa, with or without changing the colour compared to Rev1 (Table 1). The rewarding stimuli for the different phases were assigned randomly to each horse before the study. For three of the six crib-biters and four of seven control horses, there was a colour change between Rev1 and Acq2 (Table 1).

In previous studies, it was shown that extended sessions of concentrated training could lead to a lack of motivation or to inappropriate and inefficient learning behaviour (McCall 1990; Rubin et al. 1980). To ensure that horses stay motivated, we decided, based on some preliminary tests (*N* = 7 non-stereotypic horses, not used in this study) to perform two sessions of 20 trials per day, and also to reward horses in cases when they chose the incorrect flap three times in a row (and thus did not obtain any reward).

Response measures

Stereotypy level

We scored the number of crib-biting events over time from the video recordings collected over a 48 h period before the start of the experiment (see in "Experimental protocol" above), to assess the stereotypy level of the crib-biters at the time of the study. This score was converted into a frequency of crib-biting events per hour per horse. Based on these frequencies, we made three groups of crib-biters for the analyses: "S", strong crib-biters (58.37−65.76 crib-biting events per hour, *N* = 3 horses); "M", medium crib-biter (25.03 crib-biting events per hour, *N* = 1 horse); and "L" low crib-biters (1.09−9.06 crib-biting events per hour; *N* = 2 horses).

Behavioural measures

All the learning tests were video recorded using a GoPro HERO3 to control for eventual errors of scoring for all the following behaviours, which were directly scored during the tests by Experimenter 2; the choice of stimulus (correct or incorrect), the side chosen (right or left), and crib-biting events. Crib-biting was defined as instances when the horse grasped the top of the apparatus with its incisors, pulled back, contracted the neck muscles and drew air into its oesophagus, emitting an audible grunt (McGreevy et al. 1995). Three stereotypic horses did crib-bite during at least one trial, while the others never did so (Table 2).

Physiological measures

Before bringing a horse into the testing arena, it was equipped with a wireless heart-rate monitor (MLE120X Bioharness Telemetry System, Zephyr) fixed on a specific girth. During the acclimation and pre-training phases, the horses did also wear the girth for habituation, but without recording any data. During the tests, we collected the ECG trace continuously, allowing us to obtain the heart rate (HR) and the root mean square of successive inter-beat interval differences (RMSSD) as indicator of the physiological stress level of the subjects (von Borell et al. 2007). ECG gel was applied on the electrodes before each use. The data were transmitted in real time to a laptop using AcqKnowledge software v.7.2 (Biopac), and stored for later analyses. This allowed Experimenter 2 to add live comments during the visual discrimination task indicating when each session and each trial started. This enabled us to measure the physiological parameters precisely for each phase and each trial. We extracted HR and RMSSD from good-quality sections with clearly visible heartbeats on the ECG trace. We divided each session into five parts of equal duration, and analysed, when possible, three segments of 10 s each per part (at the beginning, middle, and end of each part). We checked visually that the software was tracking the heartbeats properly, and extracted HR and the inter-heartbeat (RR) intervals (ms). RR intervals were then used to calculate RMSSD (ms). We then calculated an average value per phase ("Acq1", "Rev1", "Acq2", "Rev2") for HR and for RMSSD. The total duration over which we were able to extract HR and RMSSD for the analyses was comparable between crib-biters and controls ("CB" group; means ± SD = 41.20 ± 18.59 s and "C" group = 48.28 ± 10.56 s).

Statistical analysis

All our data were analysed using generalised linear mixed models (GLMMs) or linear mixed-effect models (LMMs), in R 3.0.2, as described below. The list of the fixed and control factors as well as interaction terms included in each model are given in Tables 3 and 4. Because many studies have shown that stereotypy levels (frequency of stereotypy performance over time) can be strong predictors of learning abilities (Garner and Mason 2002; Garner et al. 2003; Vickery and Mason 2005), we first ran the models described below to investigate differences between crib-biters (“CB”) and control horses (“C”) (“GroupCB-C”), and then reran the same models again to investigate, this time, differences between the four groups defined by the frequency of crib-biting events per hour over 48 h (see “Stereotypy level” above) as follows; “S” strong crib-biters, “M” medium crib-biter, “L” low-frequency crib-biters and “C” control (“GroupSMLC”) (Tables 3, 4).

We first investigated group differences (GroupCB-C, or GroupSMLC, Tables 3, 4) in the number of training sessions the horses needed for the acclimation and pre-training phases (“Session”). The fixed and control factors as well as interactions terms included in this GLMM are described in Tables 3 and 4. The place where the horses were housed was added as a random factor. Two-by-two comparisons between the different groups of the factor GroupSMLC (“S”, “M”, “L”, and “C”) were then carried out using Tukey post-hoc tests (function glht, package multcomp in R, multiple comparisons of means).

To investigate learning performances, we then tested for group differences (GroupCB-C, or GroupSMLC, Tables 3, 4) in the frequency of correct choices (i.e. number of correct choices divided by the total number of trials; “Correct”) during the last session of acquisition for the first (“Acq1”) and second (“Acq2”) sets of learning tasks, and during the first session of reversal for the first (“Rev1”) and second (“Rev2”) sets of learning tasks. The fixed and control factors as well as interaction terms included in these LMMs are described in Tables 3 and 4. The horse identity nested within the place where the horses were housed was included as a random factor. Two-by-two comparisons between the different learning phases (between “Acq1” and “Rev1”, and between “Acq2” and “Rev2”) were then carried out using Tukey post-hoc tests (function glht, package multcomp in R, multiple comparisons of means).

Finally, to compare the learning abilities (i.e. number of trials needed to reach the learning criterion for each learning phase, “Trial”) and the physiological stress level of GroupCB-C and of GroupSMLC, we ran two separate sets of LMMs with Trial, HR or RMSSD as response variables.

Table 3 Abbreviations of the parameters used in the different analysis

Parameter abbreviation	Definition
“Phase” (c)	Different phases of learning until the learning criterion (LC) is reached
“Acq1” (c)	First acquisition: phase during which the horses learned to choose the rewarded signal among a first set of two stimuli
“Rev1” (c)	First reversal: phase during which the horses learned to choose the signal that was unrewarded in Acq1
“Acq2” (c)	Second acquisition: phase during which the horses learned to choose the rewarded signal among a novel set of stimuli
“Rev2” (c)	Second reversal: phase during which the horses learned to choose the signal that was unrewarded in Acq2
“Correct” (f)	Frequency of correct choices during the last session of Acq1 and Acq2, and during the first session of Rev1 and Rev2
“Session” (n)	Number of sessions of 10 min during the acclimation and pre-training (two sessions per day)
“Trial” (n)	Number of trials needed until the learning criterion is reached (6 correct trials in a row)
“Signal” (c)	Cross or circle
“Col” (c)	Colour of the signal (black or white)
“Person” (c)	Person leading the horse (two different persons)
GroupCB-C (“CB” or “C”) (c)	Crib-biting or control group
GroupSMLC (“S”, “M”, “L” and “C”) (c)	Groups of crib-biters based on the frequency of crib-biting events per hour over 48 h (4 groups; strong crib-biters, medium crib-biters, low-frequency crib-biters and controls)
“Colour”: “Y” or “N” (c)	Whether the colour changed between Rev1 and Acq2 (yes or no)
“HR” (m)	Heart rate (average value per horse per phase, in BPM)
“RMSSD” (m)	Root mean square of successive inter-beat interval differences (average value per horse per phase, in ms)

c category, *f* frequency, *m* mean, *n* number

Table 4 Response variables, and fixed and control factors used in the different models

	Acclimation	Learning performance	Learning capacities	Physiology
Response variable				
Session	×			
Correct		×		
Trial			×	
HR				×
RMSSD				×
Fixed factors				
GroupCB-C	×	×	×	×
GroupSMLC	×	×	×	×
Phase		×	×[a]	×[a]
Colour			×[a]	×[a]
Phase × GroupCB-C		×	×[a]	×[a]
Phase × GroupSMLC		×	×[a]	×[a]
Colour × GroupCB-C			×[a]	×[a]
Colour × GroupSMLC			×[a]	×[a]
Control factors				
Sex	×	×	×	×
Age	×	×	×	×
Person	×	×	×	×
Col		×	×	×
Signal		×		

The abbreviations are described in Table 3. The crosses indicate which factors and which response variable were used in the different models. The fixed factors are the GroupCB-C or GroupSMLC, the Phase and the Colour depending on the model. The others factors are control factors. For the model with Correct as response variable, we selected only Rev1 and Acq2 among the other Phases

[a] Indicates that we used either Phase or Colour as fixed factors in the model

The first set of LMMs was aimed at testing the learning abilities and stress levels of crib-biters and controls during each learning phase (Acq1, Rev1, Acq2 and Rev2). In this set, we investigated group differences (CB-C or SMLC) in Trial, HR or RMSSD during the four learning phases. The fixed, control factors and interaction terms included in these LMMs are described in Tables 3 and 4. Because of the small sample size, we additionally carried out a power analysis for the effect of Group CB-C and GroupSLMC on Trial, HR and RMSSD to calculate if the power of our analysis was large enough (pwr.f2 function, pwr library in R). Two-by-two comparisons between the different learning phases were then carried out using Tukey post hoc tests (function glht, package multcomp in R, multiple comparisons of means). The second set of LMMs was aimed at testing the effect of the change in the colour of the signal that some horses experienced between Rev1 and Acq2 on Trial, HR or RMSSD (Table 1). Indeed, this change in colour (hereafter “Colour”, change in colour “Y”, no change of colour “N”) could have also been perceived as a reversal by the horses. Since the factors Phase and Colour are correlated, we tested their effects on the response variables in different sets of models. In the set used to test the effect of the change in colour, the same fixed and control factors were included as in the set used to test the effect of the phase (Tables 3, 4), except for the fixed factor Phase, which was replaced by Colour. In this second set, only the data for Rev1 and Acq2 were included, because we were interested specifically in the colour change or not between Rev1 and Acq2. Because of the small sample size, we again carried out a power analysis for the effect of GroupCB-C and GroupSLMC on Trial, HR and RMSSD to calculate if the power of our analysis was large enough (pwr.f2 function, pwr library in R). For all these models (first and second sets), the horse identity nested within the place where the horses were housed was included as a random factor.

For all models described above, the residuals were checked graphically for normal distribution and homoscedasticity. To satisfy model assumptions, we used a square-root transformation for Trial, and a cube-root transformation for RMSSD. All the resulting parameters satisfying model assumptions were then entered into linear mixed-effects models fit with Gaussian family distribution and identity link function (lme function, nlme library in R). The session did not meet the statistical assumptions despite transformation. It was thus transformed to binomial data as follows; value equal or higher than median = 1 or value lower than median = 0. This parameter transformed to binomial data was input into a generalised linear mixed model fit with binomial family distribution and logit link function (glmer function, lmerTest library in R). For all models, we used a standard model simplification procedure by removing each non-significant term, until the deletion caused a reduction in goodness of fit (at which point, the term was left in the model). We assessed the statistical significance of each factor by comparing the model with and without the factor included using likelihood-ratio tests (LRT). The significance level of the factors was set at $\alpha = 0.05$.

Results

Acclimation and pre-training

Crib-biters required significantly more sessions ("Session") to fulfil the learning criterion before starting the discrimination learning task than controls (CB: 6.83 ± 2.99 sessions; C: 4.57 ± 0.79 sessions; GLMM: effect of GroupCB-C on Session; $\chi^2 = 7.29$, $df = 1$, $p = 0.007$). There was also a significant effect of the four groups defined by the frequency of crib-biting events per hour over 48 h ("GroupSLMC": "S" strong crib-biters, "M" medium crib-biter, "L" low-frequency crib-biters and "C" control) on sessions (GLMM: Effect of GroupSMLC on Session, $\chi^2 = 11.94$, $df = 4$, $p = 0.008$). However, further two-by-two comparisons did not show any significant difference between the four groups "S", "M", "L" and "C" in the number of sessions needed to fulfil the learning criterion ($p \geq 0.98$ for all).

Learning performance

All 13 animals completed the four learning phases ("Phase": "Acq1", "Rev1", "Acq2", "Rev2"). There was a significant effect of Phase on the frequency of correct choices ("Correct") in the last session of acquisition for the first and second sets of learning tasks, and the first session of reversal for the first and second sets of learning tasks (LMM: effect of Phase on Correct, $F_{3,36} = 15.51$, $p < 0.0001$; Fig. 2). Further post hoc analyses showed a significant drop in the frequency of correct choices per session between the last session of Acq1 (mean ± SD = 0.72 ± 0.18) and the first session of Rev1 (Rev1 = 0.40 ± 0.17; multiple comparisons of means; effect of Phase Acq1 versus Rev1 on Correct; $Z = -4.80$, $N = 13$, p < 0.0001; Fig. 2). There was also a

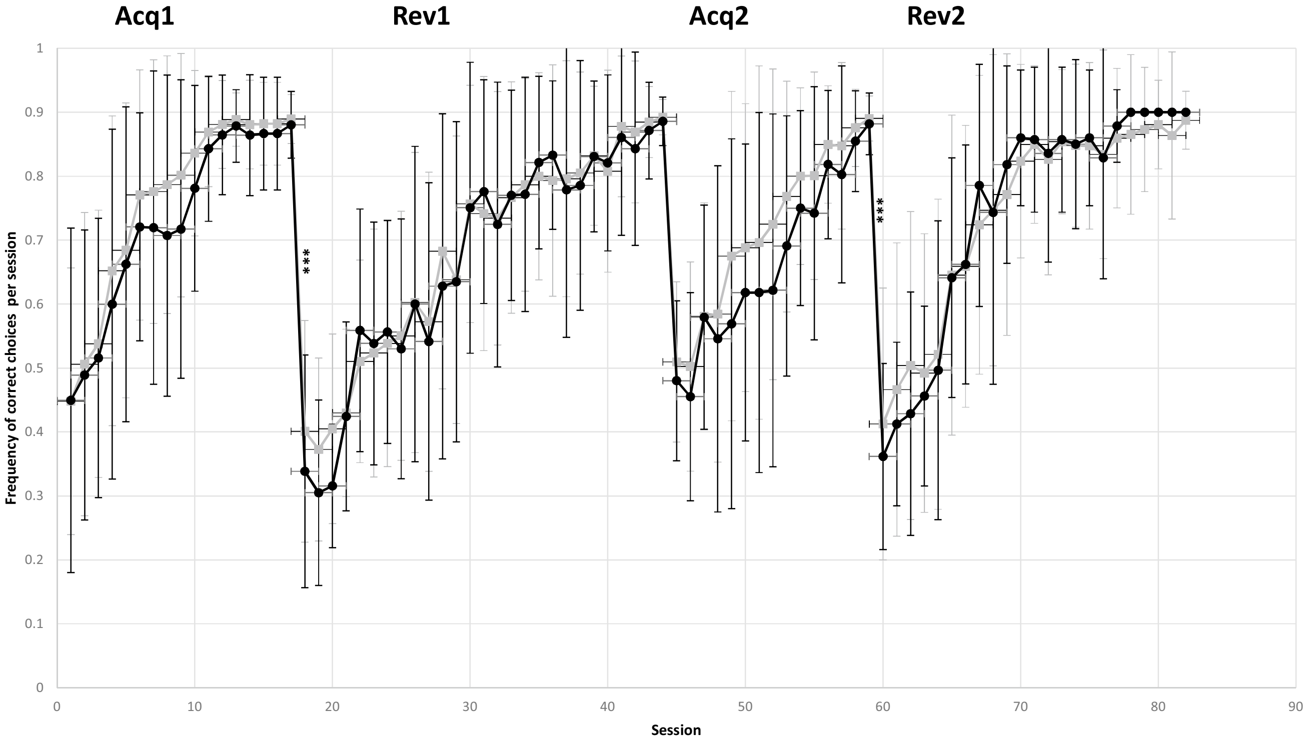


Fig. 2 Session-by-session summary of the performance of all horses. Data are the mean frequency (± SD) of correct choices per session per group (control horses in black and crib-biters in grey) combined. The dots indicate the different session. For each horse, once the learning criterion (6 correct trials in a row) was reached, it was assigned a score of 90% until all remaining animals reached the criterion within that acquisition or reversal phase. The different phases are the first acquisition (Acq1), the first reversal (Rev1), the second acquisition (Acq2) and the second reversal (Rev2). Significant differences between the last session of Acq1 and the first session of Rev1 and between the last session of Acq2 and the first session of Rev2 are indicated (***$p \leq 0.001$)

significant drop in the frequency of correct choices per session between the last session of Acq2 (0.76 ± 0.08) and the first session of Rev2 (0.44 ± 0.22; multiple comparisons of means: effect of Phase Acq2 versus Rev2 on Correct; $Z = -4.80$, $N = 13$, $p < 0.0001$, Fig. 2). On the other hand, there was no effect of group CB versus C on the frequency of correct choice (LMM: effect of GroupCB-C on Correct; $F_{1,7} = 1.77$, $p = 0.31$) nor of GroupSMLC (LMM: effect of GroupSMLC on Correct; $F_{3,5} = 0.59$, $p = 0.65$).

Learning capacities

There was neither an effect of GroupCB-C (LMM: effect of GroupCB-C on Trial; $F_{1,7} = 1.77$, $p = 0.23$; Fig. 3), nor of GroupSMLC (LMM: effect of GroupSMLC on Trial; $F_{3,5} = 0.962$, $p = 0.48$), on the number of trials per phase needed to reach the learning criterion ("Trial"). However, there was a significant difference between phases (Acq1; Rev1; Acq2; Rev2) in Trial for all horses (LMM: effect of Phase on Trial; $F_{3,36} = 5.05$, $p = 0.005$; Fig. 3). The number of trials needed until the learning criterion was reached is shown in Fig. 3. Post-hoc comparisons showed that all horses needed significantly more trials for Rev1 than for Acq1 (multiple comparisons of means: $Z = 3.64$, $N = 13$, $p = 0.002$; Fig. 3), and more trials for Rev1 than for Acq2 (multiple comparisons of means: $Z = 2.92$, $N = 13$, $p = 0.018$; Fig. 3) and for Rev2 (multiple comparisons of means: $Z = -2.65$, $N = 13$, $p = 0.041$ Fig. 3). The other two-by-two comparisons were not significant ($P \geq 0.76$ for all). In addition, considering only the first reversal and second acquisition, horses needed more trials when there was a colour change ("Colour") (mean ± SD = 169.86 ± 95.12 trials) than when there was no change in colour (100.67 ± 87.13 trials) between Rev1 and Acq2 (LMM: effect of Colour on Trials; $F_{1,12} = 6.603$, $p = 0.025$).

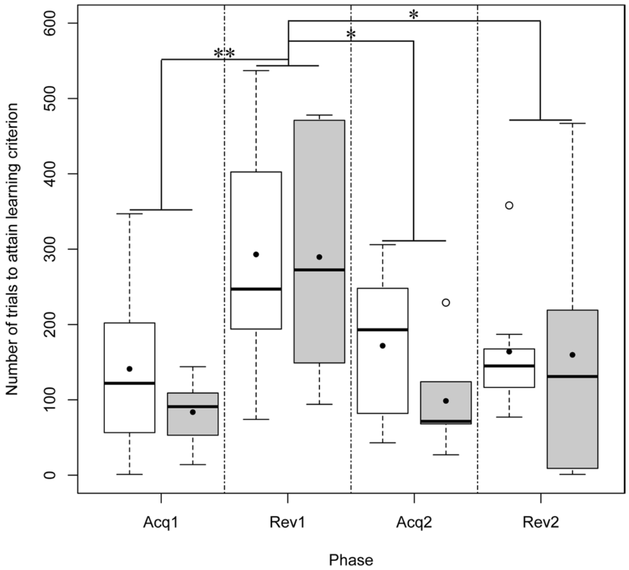


Fig. 3 Learning capacities of crib-biters and control horses. Number of trials until the learning criterion was attained for all crib-biters (group CB: $N = 6$ horses, in grey) and control horses (group C: $N = 7$ horses, in white), for each Phase (Acq1, Rev1, Acq2, Rev2). The different phases are the first acquisition (Acq1), the first reversal (Rev1), the second acquisition (Acq2) and the second reversal (Rev2). The learning criterion was fixed at six correct trials in a row. The black dots indicate the means per group per phase. Significant differences between phases are indicated as $^{*}p \leq 0.05$; $^{**}p \leq 0.001$

For all the LMMs carried out on Trial, neither the interaction between Phase and GroupCB-C or GroupSMLC, nor the sex, the age, the person leading the horse (two different persons), the type of signal (cross or circle), or the colour of the signal (black or white) had a significant effect. These terms were thus removed during model selection. Power analyses conducted on non-significant models revealed that the various LMMs testing the effect of GroupCB-C and GroupSMLC on Trial had a power ≥ 0.94, suggesting that a larger sample size would not have led to a significant result.

Physiological parameters

Heart rate (HR)

There was no difference in HR between the different learning phases (LMM: effect of Phase on HR; $F_{3,35} = 1.03$, $p = 0.39$), no effect of GroupCB-C (LMM: effect of GroupCB-C on HR; $F_{1,7} = 2.09$, $p = 0.20$), and no effect of GroupSMLC (LMM: effect of GroupSMLC on HR; $F_{3,5} = 1.88$, $p = 0.25$). In addition, considering only the first reversal and second acquisition, Colour had no effect on HR (LMM: effect of Colour on HR; $F_{1,12} = 0.30$, $p = 0.59$).

Root-mean square of successive inter-beat interval differences (RMSSD)

There was no difference in RMSSD between the different learning phases (LMM: effect of Phase on RMSSD; $F_{3,32} = 0.97$, $p = 0.41$), no effect of GroupCB-C (LMM: effect of GroupCB-C on RMSSD; $F_{1,7} = 0.05$, $p = 0.84$), and no effect of GroupSMLC on RMSSD (LMM: effect of GroupSMLC on RMSSD; $F_{3,5} = 0.91$, $p = 0.50$). In addition, Colour had no effect on RMSSD (LMM: effect of Colour on RMSSD; $F_{1,12} = 1.35$, $p = 0.27$).

For all the LMMs carried out on HR or RMSSD, neither the interaction between Phase and Group, nor the interaction between Colour and Group, nor the sex, the age, the person leading the horse (two different persons), the signal (cross or circle), or the colour of the signal had a significant effect. These terms were thus removed during model selection. Power analyses conducted on non-significant models revealed that the various LMMs testing the effect of GroupCB-C and GroupSLMC on HR or RMSSD had a power ≥ 0.87, suggesting that a larger sample size would not have led to a significant result.

Discussion

In this study, we used a reversal learning task, which has been used as a diagnostic for basal ganglia dysfunction, to compare the learning performances of crib-biting and control horses. According to our results, there is no indication that crib-biters suffer from such a dysfunction. Except for the acclimation phases, which took longer for crib-biters compared to the controls to achieve, we did not find any differences between crib-biters and control horses in the number of trials necessary to reach the learning criterion in any phase of the experiment. In fact, all horses reached the learning criterion and performed the two reversals. Interestingly, they also performed the second reversal in fewer trials compared to the first one, suggesting that they learned to learn. Unlike in other studies that found that crib-biting horses have altered learning abilities compared to other horses (Hemmings et al. 2007; Parker et al. 2008, 2009; Roberts et al. 2015), our subjects had the opportunity to crib-bite during the experiment, and hence to potentially reduce their stress levels (Briefer Freymond et al. 2015), as shown by the absence of differences between the two groups in the physiological parameters that we measured. We could therefore suggest that discrepancies between our study and previous studies on learning performance could result from differences in stress levels experienced by crib-biters and control horses in these former studies, although they did not collect physiological measures of stress.

Acclimation and pre-training

In this study, crib-biters needed a longer time than the controls to be acclimated to the learning apparatus, and to attain the conditions required to start the discrimination procedure (i.e., pushing the flaps alone five times on both sides of the apparatus without any intervention of the experimenter). This could be explained by the fact that crib-biters seem to be more stress sensitive (Briefer Freymond et al. 2015), and might thus need more time to be acclimated to a new situation. However, we did not collect any physiological indicators of stress during the pre-training phase, because in this phase, the movement of the horses was not standardised, unlike during the learning phase. Indeed, it is recommended that only measures made during times of similar behavioural pattern should be compared (von Borell et al. 2007). Future studies that additionally measure stress parameters during habituation could inform us on the stress levels of crib-biters in such situations.

Learning capacities

All the horses (crib-biters and controls) in our study performed the two reversal tasks, and needed significantly more trials to reach the learning criterion for the first reversal (“Rev1”) than for the other phases (first acquisition, “Acq1”, second acquisition, “Acq2”, and second reversal, “Rev2”). In addition, the significant drop that we observed in the number of correct responses between the acquisition phases and their following respective reversals suggests that learning had taken place (McBride and Parker 2015). Although few studies suggest that horses possess the ability to perform reversal learning based only on visual cues (Sappington et al. 1997; Voith 1975), this task seems to be more challenging, and in some cases not achievable, compared to reversal learning based on spatial cues (Hothersall et al. 2010; Martin et al. 2006). The fact that control and even stereotypic horses learned the reversal task let us suggest that reversal learning based only on visual cue is possible under certain conditions. Moreover, the fact that horses performed the second reversal in less trials than the first one, confirmed that horses learned to learn as it was also demonstrated in other studies testing horses in visual or spatial discrimination tasks (Fiske and Potter 1979; Martin et al. 2006; Voith 1975; Warren and Warren 1962). As underlined by Brubaker and Udell (2016), the study protocol and nature of the visual stimuli appear to affect a horse’s ability to perform any given cognitive task. In our study, we adapted the experimental protocol, based on preliminary tests and previous studies (Flannery 1997; Gabor and Gerken 2010; Hall et al. 2003) to keep the horses motivated, as follows. First, we chose to oppose two signals that differed only in whether they were black or white, because such colours seem to be easy to differentiate by horses. Indeed, horses’ ability to discriminate between different colours is limited due to their dichromatic vision (Blackmore et al. 2008). Second, we ensured that the stimuli were presented close to the ground level (40 cm in our study). Indeed, former studies demonstrated that horse performance was improved when stimuli were presented close to the ground level (22 cm), compared at the height of 90 cm from the ground (Hall et al. 2003). Third, we adapted the number of trials per sessions and rewarded the horses during the study if they had been choosing the wrong stimulus three times in a row, to maintain their attention span and motivation (Flannery 1997; Rubin et al. 1980; Sappington et al. 1997). Finally, we waited until all the horses reached the learning criterion before stopping the study, even when a relatively high number of trials was required (N max = 537). We suggest that similar precautions might help improve the motivation of horses in future cognitive studies.

Learning performance of crib-biters compared to control horses

We did not find any difference in the number of trials needed to reach the learning criterion between crib-biter and control horses, contrary to previous studies on the same topic (Hemmings et al. 2007; Parker et al. 2008, 2009; Roberts et al. 2015). Indeed, previous studies found that crib-biters might be more prone to habit learning than to response–outcome learning (Hemmings et al. 2007; Parker et al. 2008, 2009; Roberts et al. 2015). Therefore, we expected that they would reach the learning criterion during Acq1 or Acq2 faster compared to the controls. In addition, since previous studies also demonstrated that crib-biting horses need more operant responses compared to the other horses before the extinction of a previously learnt action (Hemmings et al. 2007; Roberts et al. 2015), and were unable to maintain response–outcome learning in a continuously applied learning paradigm (Parker et al. 2008), we expected that the crib-biters would need more trials in Rev 1 and Rev 2 to reach the criterion compared to controls. In contrast to these predictions, in our study, all the crib-biters were able to achieve the different phases (Acq1, Rev1, Acq2, Rev2) in a similar number of trials compared to the controls. An explanation for these discrepancies between our studies and previous ones could be that in our experiment, the crib-biters had the opportunity to crib-bite on the learning apparatus. It is not always clear whether stereotypic horses had the opportunity to crib-bite, and did so, during previous studies (Hemmings et al. 2007; Parker et al. 2008, 2009). However, in Roberts et al. (2015), crib-biting straps were removed prior to the tests, although no information about crib-biting events is specified. If crib-biting is indeed a coping strategy (Briefer Freymond et al. 2015), reducing stress levels could, as a result, improve their learning capacities, allowing them to achieve the same performances as non-stereotypic horses (Schwabe and Wolf 2010; Valenchon et al. 2013). Even if other studies did not measures stress parameters during learning tasks (Hemmings et al. 2007; Parker et al. 2009, 2008; Roberts et al. 2015), this suggests that allowing crib-biting horses to perform their stereotypic behaviour during learning could improve their cognitive abilities.

Since stereotypies in animals are often likened to human developmental, neurological or severe psychiatric disorders (e.g., autism, obsessive compulsive disorder (OCD) or schizophrenia) (McBride and Parker 2015), comparisons between our results and human disorders can be made. Even if OCD patients usually report that they get a sort of relief by performing their rituals, and that preventing performance increases their anxiety (Boyer and Lienard 2006), the literature on autism in humans suggests that, on the contrary, children exhibiting high levels of stereotypy fail to learn while engaged in stereotypy (Cunningham and Schreibman 2008). On the other hand, our findings could be related to results found in another human psychological disorder, named attention deficit hyperactivity disorder (ADHD). In the same way as animals suffering from stereotypies, ADHD patients show non-goal-oriented motor movements. These movements are however, unlike stereotypies, not executed as repetitive invariant patterns. A recent paper showed that performing such movements is associated with an improvement in cognitive performance (Sarver et al. 2015). Such findings are in accordance with our results in crib-biting horses. As suggested in Hausberger et al. (2007), stereotypic horses differ from other horses in their behaviour and may require specific training. Letting these horses the possibility to perform their stereotypy might be, as our study suggests, one specific feature to incorporate in learning protocols, which might then allow them to perform successfully. Future studies could compare the learning capacities of crib-biter horses prevented or not to crib-bite and of a corresponding number of control horses subjected to the same treatment. To summarize, the results of our study do not support the hypothesis that crib-biters display alterations in learning abilities, which could result from impaired dopaminergic system. In addition, our findings suggest that, in the same way as the performance of non-goal-oriented motor movements improves cognitive performances of ADHD human patients (Sarver et al. 2015), the performance of stereotypic behaviour might improve crib-biting horse learning abilities.

During the acclimation and pre-training phase, however, crib-biters needed more sessions than controls to attain the learning criterion, even if they also had the possibility to crib-bite. Since we did not collect measures of stress indicators (e.g., HR, RMSSD) during this period, we cannot make inferences about the stress level of stereotypic horses compared to controls during this phase. Nevertheless, we could suggest that crib-biting horses were less focussed on the task during the acclimation and pre-training compared to the test phase, because horses had to manage too many other external stimuli. A general difficulty of maintaining a task focus (i.e., attention) in stereotypic compared to control horses has been suggested by Hausberger et al. (2007). These authors proposed that the time invested in performing stereotypic behaviour throughout the day and at night is likely to affect sleep quality and quantity in stereotypic horses. This could lead to a general lower attention span in these horses than in non-stereotypic horses. Attention state and motivation being primordial for learning (Cowan 1998; Rochais et al. 2014), we could hypothesise that crib-biters might needed a longer time than control horses to habituate to novel situations and be able to ignore and exclude external stimuli (i.e., new area, apparatus). However, after a longer habituation than for the controls, they might be able to focus on the cognitive task (i.e., test phase). During the test phase itself, as a result of our protocol that was designed to maximise attention span and motivation (e.g., short training sessions of around 20 min, "correction trials" (Flannery 1997)), all horses, including crib-biters, seemed very attentive and motivated, which might have boosted their performances. Attention deficits have also been demonstrated in ADHD disorder, which has the particularity to induce difficulties in maintaining task focus (Sarver et al. 2015). To test the hypothesis that crib-biters are generally less attentive than other horses, future studies could evaluate the distractibility (i.e., how much can an individual be distracted by external stimuli) of these horses compared to non-stereotypic individuals (e.g., using a "distractibility test"; (Rochais et al. 2017)). Another indicator of attention that could be used to test such hypothesis is spontaneous blink rate (SBR; (Roberts et al. 2015)). SBR is a basic measure of dopamine transmission utilised to determine striatal functioning in stereotypy-performing humans and could also be applied to animals (Roberts et al. 2015). Using this indicator, Roberts et al. (2015) demonstrated that crib-biter horses display lower SBR than other horses (Roberts et al. 2015). Interestingly, SBR has been recently shown to constitute an index of dopaminergic component of sustained attention and fatigue in humans (Maffei and Angrilli 2018). By comparing the distractibility and attention of stereotypic and control horses when performing cognitive tests, future studies might be able to further highlight the need to adapt the design of training sessions to horses suffering from stereotypies, in order to maximum their learning abilities and improve their welfare (e.g., by avoiding frustration).

A last explanation for the discrepancies between our studies and other studies investigating learning alteration in crib-biting horse could be that reversal learning tests are perhaps not appropriate for assessing dopaminergic alterations that might be present in crib-biters. Such alterations, including a higher number of dopamine receptor subtypes in the ventral striatum or nucleus accumbens *(Nac)* and a lower number of such receptors in the dorsomedial striatum (DMS) or caudate have been shown using post-mortem analyses in crib-biting horses (McBride and Hemmings 2005). However, behavioural flexibility, or the ability to adjust responses according to a change in the environment, is mediated by a large neural network, including prefrontal-basal ganglia circuits in addition to the dorsal and ventral striatum. As mentioned previously, the DMS has been identified as an important structure for flexible responding (Castane et al. 2010). Indeed, DMS lesions, due to the role of this brain structure in learning, might result in a switch from goal-directed to habit formation and thus in the impairment of the development of habits (Yin et al. 2008). However, the role of *Nac* in instrumental performance remains nowadays controversial (Yin et al. 2008). For example, some studies found that lesions in the *Nac* do not impair spatial, visual or motor reversal in monkeys, *Macaca fascicularis* (Stern and Passingham 1995), while other studies found that such lesions impaired both an initial discrimination and its reversal in Lister hooded rats (Annett et al. 1989). A reason for these discrepancies between studies could be that most studies on DMS or dorsolateral (DLS) lesions have used rats, despite the fact that it is difficult to compare the physical location of dorsal or ventral striatum in rat and other species such as primates (Yin et al. 2008). To conclude, reversal learning paradigms are among the most widely used tests for cognitive flexibility and there is accumulating evidence that DMS is involved in this type of learning (Castane et al. 2010; Izquierdo et al. 2017; Ragozzino et al. 2003). However the role of *Nac*, which has been suggested to be impaired in crib-biters (McBride and Hemmings 2005), in reversal learning tasks is controversial (Yin et al. 2008). Therefore, testing crib-biters with other cognitive tests than reversal learning might be perhaps more valuable to investigate the suggested impairment in *Nac* (McBride and Hemmings 2005).

Effect of the change in colour

In our study, the colour of the signal always changed from white to black or vice versa, between the acquisition and the corresponding reversal. However, for half of the horses (half of the crib-biters and four of the seven controls), a change in colour occurred also in the middle of the learning procedure, between Rev1 and Acq2 (Table 1; Fig. 1). This change in colour in the middle might have been experienced as an additional reversal (based on colour only). In accordance with this hypothesis, the results showed a significant effect of colour change between Rev1 and Acq2, with horses submitted to the colour change needing more trials than the other horses. We suggest that further studies including several acquisition phases with different visual stimuli should be aware that changes in colours between phases might be perceived by the animals as reversals.

Stereotypy level and performance

The crib-biters in this study differed with regard to the strength of their stereotypy (Table 1). Many studies have reported more cognitive difficulties in animals displaying a higher frequency of stereotypic behaviour compared to less stereotypic ones (Garner and Mason 2002; Garner et al. 2003; Vickery and Mason 2005). Indeed, stereotypic levels have been shown to correlate with an increase in the persistence of inappropriate responses in an extinction learning test in bears, *Ursus thibetanus* (Vickery and Mason 2005). However, our results did not show such a trend. Our three groups based on the frequency of crib-biting of the horses assessed over 48 h before the experiment started (GroupSMLC, "S", strong crib-biters, "M", medium crib-biters and "L", low crib-biters), did not differ in their learning performance. Therefore, we did not find any evidence showing that the frequency of crib-biting is a factor that influences the cognitive abilities of horses. This absence of group difference could also suggest that the stereotypic level is not a good indicator of dopaminergic system alterations. Similar results have been found in rhesus macaques, *Macaca mulatta* (Pomerantz et al. 2012). Interestingly, in this study, the authors found that some type of stereotypies did correlate with perseveration while some did not. Future studies could investigate the potential links between learning performance and the time since a horse started crib-biting instead of its stereotypic level.

Physiological parameters

We did not find any evidence for group differences in the sympathomedullary (SAM) axis parameters measured in this study (HR and RMSSD) during the phases of acquisitions and their respective reversals. Because one potential cause of stereotypies is a previous exposure to a chronic stress situation that could induce higher sensitivity to stress (Bhatnagar and Vining 2003), we would have expected crib-biters to be more stressed than controls. However, neither the results of the present study, nor those of our previous study revealed any difference in SAM parameters between crib-biters and controls (Briefer Freymond et al. 2015). It also suggests that the crib-biters were experiencing similar stress levels as controls during the experiment, possibly as a result of crib-biting on the apparatus, which might have reduced their stress levels (Briefer Freymond et al. 2015). We could also have expected horses to be more physiologically stressed during the first reversal compared to the other learning phase as this learning phase might be more challenging for them, as displayed by the increased number of trials required to achieve this task. However, we did not find any effect of the learning phase (acquisition or reversal) on HR or RMSSD. Finally, it is possible that other parameters than HR and RMSSD might be more adequate to measure stress during a learning task involving locomotor behaviour. Indeed, HR and RMSSD are also influenced by physical activity (von Borell et al. 2007). For this reason, only measures made during times of similar behavioural pattern should be compared (von Borell et al. 2007). Our assumption is that this is the case in our study, because all horses (crib-biters and controls) had to perform the same trajectory, and the same number of trials per session. However, further studies could aim at designing tasks involving less movement and take also additional measures of stress, such as behavioural measures (e.g. ‚Equine Facial Action Coding Systems (FACS) (Wathan et al. 2015), behaviour scores (Young et al. 2012)) during learning tasks as well as during habituation.

Animal welfare

Animal welfare being of increasing public and scientific concern, it is important to understand the link between stereotypic behaviour and animal welfare. It has been proposed that stereotypic behaviour might indicate poor welfare only if a dopaminergic dysfunction is present (Mason and Latham 2004). Indeed, even if stereotypies develop under contexts of chronic stress state, their performance, once fully developed, might not necessarily indicate poor welfare. For instance, in this case, the performance of stereotypic behaviours might serve as coping mechanisms, helping individuals to reduce their stress level (e.g., *"mantra effect"*, (Mason and Latham 2004)). Mason and Latham (2004) proposed that stereotypies correlate with poor welfare only when they have become a habit and, only when behaviours have changed in control and have become environmentally insensitive. At this developmental stage, stereotypies are performed in a more diverse set of situations and are harder to interrupt. However, such a stage is not easy to assess because the performance of stereotypies may vary between individuals, in terms of stereotypy level over time. Mason and Latham (2004) additionally argued that perseverative responding, resulting from basal ganglia dysfunction, also indicates poor welfare. In fact, with perseveration, individuals may produce unnecessary and inappropriate responses to environmental cues. In humans, as mentioned earlier, perseveration is also correlated with human disorders, like schizophrenia, autism and other brain injuries. In conclusion, if neurobiological changes are linked to stereotypies, resulting in alterations in the learning profile of animals, stereotypic behaviour should indicate poor welfare. However, in this study, we could not conclude to the existence neurobiological alteration in crib-biters, since these horses did not need more trials to perform the reversal learning tasks compared to control horses. Further studies on cognitive abilities of crib-biter horses are thus required to determine the impact of this stereotypy on horse welfare.

Conclusion

Our study did not reveal any difference in cognitive abilities between crib-biters and controls and therefore we cannot conclude that stereotypic horses suffer from a dopaminergic dysfunction. Indeed, our results show that all horses, including stereotypic horses and controls, were able to perform reversal discrimination tasks based on visual cues, and that they even learned to learn (i.e., improve their performance from one reversal to the next). An explanation for the discrepancies between our study and the previous ones could be that, in our study, the crib-biters had the opportunity to crib-bite on the learning apparatus, which might have enabled them to reduce their stress level, as suggested by the lack of group difference in physiological stress parameters. Further studies could test the learning capacities of crib-biters that are prevented or not to perform the stereotypic behaviour against a group of non-stereotypic horses subjected to the same treatment. Finally, our results point towards several parallels between horse crib-biting behaviour and human developmental, neurological or psychiatric disorders, such as ADHD disorders, suggesting that the study of horse crib-biting behaviour could serve as a good animal model to better understand such disorders in humans (Brace et al. 2015).

Acknowledgements We are grateful to all the owners of the horses who offered to participate in this study. We are also grateful to Matthew Parker, who gave us the idea to perform this study and helped us with the design of the experiment.

Compliance with ethical standards

Ethical approval The experimental procedure for the horses was approved by the Federal Veterinary Office (approval number VD 26777 bis; Switzerland).

Conflict of interest The authors declare that they have no competing interest.

References

Annett LE, McGregor A, Robbins TW (1989) The effects of ibotenic acid lesions of the nucleus accumbens on spatial learning and extinction in the rat. Behav Brain Res 31:231–242

Bhatnagar S, Vining C (2003) Facilitation of hypothalamic pituitary adrenal responses to novel stress following repeated social stress using the resident/intruder paradigm. Hormon Behav 43:158–165

Blackmore TL, Foster TM, Sumpter CE, Temple W (2008) An investigation of colour discrimination with horses (*Equus caballus*). Behav Proc 78:387–396

Boyd BA, McBee M, Holtzclaw T, Baranek GT, Bodfish JW (2009) Relationships among repetitive behaviors, sensory features, and executive functions in high functioning autism. Res Autism Spectr Disord 3:959–966. https://doi.org/10.1016/j.rasd.2009.05.003

Boyer P, Lienard P (2006) Why ritualized behavior? Precaution Systems and action parsing in developmental, pathological and cultural rituals. Behav Brain Sci 29:595–613; (**discussion 613–550**)

Brace LR, Kraev I, Rostron CL, Stewart MG, Overton PG, Dommett EJ (2015) Auditory responses in a rodent model of attention deficit hyperactivity disorder. Brain Res 1629:10–25

Briefer Freymond S, Bardou D, Briefer EF, Bruckmaier R, Fouché N, Fleury J, Maigrot AL, Ramseyer A, Zuberbühler K, Bachmann I (2015) The physiological consequences of crib-biting in horses in response to an ACTH challenge test. Physiol Behav 151:121–128. https://doi.org/10.1016/j.physbeh.2015.07.015

Brubaker L, Udell MAR (2016) Cognition and learning in horses (*Equus caballus*): what we know and why we should ask more. Behav Proc 126:121–131. https://doi.org/10.1016/j.beproc.2016.03.017

Cabib S, Giardino L, Calzà L, Zanni M, Mele A, Puglisi-Allegra S (1998) Stress promotes major changes in dopamine receptor densities within the mesoaccumbens and nigrostriatal systems. Neuroscience 84:193–200. https://doi.org/10.1016/S0306-4522(97)00468-5

Castane A, Theobald DE, Robbins TW (2010) Selective lesions of the dorsomedial striatum impair serial spatial reversal learning in rats. Behav Brain Res 210:74–83. https://doi.org/10.1016/j.bbr.2010.02.017

Cowan N (1998) Attention and memory: an integrated framework. Oxford University Press, Oxford

Cunningham AB, Schreibman L (2008) Stereotypy in Autism: The Importance of Function. Res Autism Spectr Disord 2:469–479. https://doi.org/10.1016/j.rasd.2007.09.006

Dickinson A (1985) Actions and habits: the development of behavioural autonomy. Philos Trans R Soc Lond B Biol Sci 308:67–78. https://doi.org/10.1098/rstb.1985.0010

Fiske JC, Potter GD (1979) Discrimination Reversal Learning in Yearling Horses. J of Anim Sci 49:583–588. https://doi.org/10.2527/jas1979.492583x

Flannery B (1997) Relational discrimination learning in horses. Appl Anim Behav Sci 54:267–280. https://doi.org/10.1016/S0168-1591(97)00006-3

Gabor V, Gerken M (2010) Horses use procedural learning rather than conceptual learning to solve matching to sample. Appl Anim Behav Sci 126:119–124. https://doi.org/10.1016/j.applanim.2010.06.006

Garner JP, Mason GJ (2002) Evidence for a relationship between cage stereotypies and behavioural disinhibition in laboratory rodents. Behav Brain Res 136:83–92. https://doi.org/10.1016/S0166-4328(02)00111-0

Garner JP, Meehan CL, Mench JA (2003) Stereotypies in caged parrots, schizophrenia and autism: evidence for a common mechanism. Behav Brain Res 145:125–134

Hall CA, Cassaday HJ, Derrington AM (2003) The effect of stimulus height on visual discrimination in horses. J Anim Sci 81:1715–1720. https://doi.org/10.2527/2003.8171715x

Hausberger M, Gautier E, Müller C, Jego P (2007) Lower learning abilities in stereotypic horses. Appl Anim Behav Sci 107:299–306. https://doi.org/10.1016/j.applanim.2006.10.003

Hemmings A, McBride SD, Hale CE (2007) Perseverative responding and the aetiology of equine oral stereotypy. Appl Anim Behav Sci 104:143–150. https://doi.org/10.1016/j.applanim.2006.04.031

Hothersall B, Gale E, Harris P, Nicol C (2010) Cue use by foals (Equus caballus) in a discrimination learning task. Anim Cogn 13:63–74

Houpt KAM, McDonnell SM (1993) Equine stereotypies. Compend Contin Educ 15:1265–1271

Hughes BO, Duncan IJH (1988) The notion of ethological 'need', models of motivation and animal welfare. Anim Behav 36:1696–1707. https://doi.org/10.1016/S0003-3472(88)80110-6

Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A (2017) The neural basis of reversal learning: An updated perspective. Neuroscience 345:12–26. https://doi.org/10.1016/j.neuroscience.2016.03.021

Lopez BR, Lincoln AJ, Ozonoff S, Lai Z (2005) Examining the relationship between executive functions and restricted, repetitive symptoms of Autistic Disorder. J Autism Dev Disord 35:445–460. https://doi.org/10.1007/s10803-005-5035-x

Luescher UA, McKeown DB, Halip J (1991) A cross-sectional study on compulsive behavior (stable vices) in horses. Equine Vet J 27:14–18

Maffei A, Angrilli A (2018) Spontaneous eye blink rate: An index of dopaminergic component of sustained attention and fatigue. Interna J Psychophysiol 123:58–63. https://doi.org/10.1016/j.ijpsycho.2017.11.009

Martin TI, Zentall TR, Lawrence L (2006) Simple discrimination reversals in the domestic horse (*Equus caballus*): effect of discriminative stimulus modality on learning to learn. Appl Anim Behav Sci 101:328–338. https://doi.org/10.1016/j.applanim.2006.02.011

Mason G, Latham N (2004) Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? Anim Welf 13:57–69

Mason G, Rushen J (2006) Stereotypic animal behaviour: fundamentals and applications to welfare, 2nd edn. CAB Interantional, Wallingford

McBride SD, Hemmings A (2005) Altered mesoaccumbens and nigro-striatal dopamine physiology is associated with stereotypy development in a non-rodent species. Behav Brain Res 159:113–118. https://doi.org/10.1016/j.bbr.2004.10.014

McBride SD, Parker MO (2015) The disrupted basal ganglia and behavioural control: An integrative cross-domain perspective of spontaneous stereotypy. Behav Brain Res 276:45–58. https://doi.org/10.1016/j.bbr.2014.05.057

McBride SD, Perentos N, Morton AJ (2016) A mobile, high-throughput semi-automated system for testing cognition in large non-primate animal models of Huntington disease. J Neurosci Methods 265:25–33. https://doi.org/10.1016/j.jneumeth.2015.08.025

McBride SD, Parker MO, Roberts K, Hemmings A (2017) Applied neurophysiology of the horse; implications for training, husbandry and welfare. Appl Anim Behav Sci 190:90–101. https://doi.org/10.1016/j.applanim.2017.02.014

McCall CA (1990) A review of learning behavior in horses and its application in horse training. J Anim Sci 68:75–81. https://doi.org/10.1093/ansci/68.1.75

McGreevy P, McLean A (2010) Equitation science. Wiley, Hoboken, p 328

McGreevy PD, Cripps PJ, French NP, Green LE, Nicol CJ (1995) Management factors associated with stereotypic and redirected behaviour in the Thoroughbred horse. Equine Vet J 27:86–91

Parker M, Redhead ES, Goodwin D, McBride SD (2008) Impaired instrumental choice in crib-biting horses (Equus caballus). Behav Brain Res 191:137–140. https://doi.org/10.1016/j.bbr.2008.03.009

Parker M, McBride SD, Redhead ES, Goodwin D (2009) Differential place and response learning in horses displaying an oral stereotypy. Behav Brain Res 200:100–105. https://doi.org/10.1016/j.bbr.2008.12.033

Pomerantz O, Paukner A, Terkel J (2012) Some stereotypic behaviors in rhesus macaques (Macaca mulatta) are correlated with both perseveration and the ability to cope with acute stressors. Behav Brain Res 230:274–280. https://doi.org/10.1016/j.bbr.2012.02.019

Ragozzino ME, Kim J, Hassert D, Minniti N, Kiang C (2003) The contribution of the rat prelimbic-infralimbic areas to different forms of task switching. Behav Neurosci 117:1054–1065. https://doi.org/10.1037/0735-7044.117.5.1054

Ringhofer M, Yamamoto S (2017) Domestic horses send signals to humans when they face with an unsolvable task. Anim Cogn 20:397–405. https://doi.org/10.1007/s10071-016-1056-4

Roberts K, Hemmings A, Moore-Colyer M, Hale CE (2015) Cognitive differences in horses performing locomotor versus oral stereotypic behaviour. Appl Anim Behav Sci 168:37–44. https://doi.org/10.1016/j.applanim.2015.04.015

Roberts K, Hemmings AJ, McBride SD, Parker MO (2017) Causal factors of oral versus locomotor stereotypy in the horse. J Vet Behav Clin Appl Res 20:37–43. https://doi.org/10.1016/j.jveb.2017.05.003

Rochais C, Henry S, Sankey C, Nassur F, Gorecka-Bruzda A, Hausberger M (2014) Visual attention, an indicator of human-animal relationships? A study of domestic horses (*Equus caballus*). Front in Psychol 5:108

Rochais C, Henry S, Hausberger M (2017) Spontaneous attention-capture by auditory distractors as predictor of distractibility: a study of domestic horses (Equus caballus). Sci Rep 7:15283

Rubin L, Oppegard C, Hindz HF (1980) The effect of varying the temporal distribution of conditioning trials on equine learning behavior. J of Anim Sci 50:1184–1187

Sappington BKF, McCall CA, Coleman DA, Kuhlers DL, Lishak RS (1997) A preliminary study of the relationship between discrimination reversal learning and performance tasks in yearling and 2-year-old horses. Appl Anim Behav Sci 53:157–166. https://doi.org/10.1016/S0168-1591(96)01157-4

Sarver DE, Rapport MD, Kofler MJ, Raiker JS, Friedman LM (2015) Hyperactivity in attention-deficit/hyperactivity disorder (ADHD): impairing deficit or compensatory behavior? J Abnorm Child Psychol 43:1219–1232. https://doi.org/10.1007/s10802-015-0011-1

Schwabe L, Wolf OT (2010) Learning under stress impairs memory formation. Neurobiol Learn Mem 93:183–188. https://doi.org/10.1016/j.nlm.2009.09.009

Schwabe L, Wolf OT (2011) Stress-induced modulation of instrumental behavior: From goal-directed to habitual control of action. Behav Brain Res 219:321–328. https://doi.org/10.1016/j.bbr.2010.12.038

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

Stern CE, Passingham RE (1995) The nucleus accumbens in monkeys (*Macaca fascicularis*). III. Reversal learning. Exp Brain Res 106:239–247

Toates F (2004) Cognition, motivation, emotion and action: a dynamic and vulnerable interdependence. Appl Anim Behav Sci 86:173–204. https://doi.org/10.1016/j.applanim.2004.02.010

Valenchon M, Lévy F, Fortin M, Leterrier C, Lansade L (2013) Stress and temperament affect working memory performance for disappearing food in horses, *Equus caballus*. Anim Behav 86:1233–1240. https://doi.org/10.1016/j.anbehav.2013.09.026

Van der Harst JE, Baars AM, Spruijt BM (2003) Standard housed rats are more sensitive to rewards than enriched housed rats as reflected by their anticipatory behaviour. Behav Brain Res 142:151–156

Vickery SS, Mason GJ (2005) Stereotypy and perseverative responding in caged bears: further data and analyses. Appl Anim Behav Sci 91:247–260

Voith VL (1975) Pattern discrimination, learning set formation, memory retention, spatial and visual reversal learning by the horse. The Ohio State University, Columbus

von Borell E 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. In: Paper presented at the physiology & behavior, stress and welfare in farm animals, 10/22/2007

Warren J, Warren HB (1962) Reversal learning by horse and raccoon. J Gen Psychol 100:215–220

Wathan J, Burrows AM, Waller BM, McComb K (2015) EquiFACS: the equine facial action coding system. PLoS one 10:e0131738

Wickens CL, Heleski CR (2010) Crib-biting behavior in horses: a review. Appl Anim Behav Sci 128:1–9

Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. Nat Rev Neurosci 7:464–476. https://doi.org/10.1038/nrn1919

Yin HH, Ostlund SB, Balleine BW (2008) Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. Eur J Neurosci 28:1437–1448. https://doi.org/10.1111/j.1460-9568.2008.06422.x

Young T, Creighton E, Smith T, Hosie C (2012) A novel scale of behavioural indicators of stress for use with domestic horses. Appl Anim Behav Sci 140:33–43