Journal of Veterinary Behavior 29 (2019) 111–117

Contents lists available at ScienceDirect

Journal of Veterinary Behavior

journal homepage: www.journalvetbehavior.com

Equine Research

Effect of preceding activity and multiple testing of Belgian police horses on behavior during personality tests

Marc Pierard [a,*], Paul McGreevy [b], Rony Geers [a]

[a] KU Leuven Department of Biosystems, Leuven, Belgium
[b] School of Veterinary Science, Faculty of Science, University of Sydney, New South Wales, Australia

ARTICLE INFO

Article history:
Received 13 March 2018
Received in revised form
18 May 2018
Accepted 9 July 2018
Available online 18 July 2018

Keywords:
police horse
personality
behavioral assessment
preceding activity
multiple testing

ABSTRACT

Operational police horses must perform under challenging circumstances, so the selection of suitable equine training candidates should include an evaluation of their personalities. We describe the developmental phase of a project aimed at customizing selection tests for mounts at the Belgian Federal Police. The experiment evaluated the influence of preceding activity and multiple testing on behavior during personality tests. Using a crossover design, three different preceding activities were compared: straight out of their box, after 30 minutes tied outside alone, and after 30 minutes alone in a paddock. A sample of 30 active police horses was tested, in three equal groups with a different treatment order per group. The personality tests included an arena and a sudden object test, measuring behavioral responses. For the sudden object test, we also recorded distance from a descending umbrella. Likert scores of 1 to 5 were used to score the ease of handling when leading the horse into and out of the arena. Order and treatment effects were analyzed with mixed models. Treatment effects were found in the arena test for 10 of 17 variables, including all measures of stand, canter and roll, and buck/jump rate. Order effects were present for stand rate in the arena and sudden object test, for rate and mean bout length of walk in the sudden object test, and for the scores of release at entering the arena and approach at the end of the testing. After all horses had been tested individually, 10 horses from the current sample were tested as duos, after coming straight out of their boxes (stables) to evaluate the effect of solitary testing and the influence of equine company as it prevails in the workplace. A Wilcoxon signed rank test that compared the duo results to those of the individual tests for the same treatment, showed a significant difference in only two of 52 variables (stand mean bout length in the arena test and Likert score for entering the arena and stopping). Conversely, only five variables had significant Spearman correlations between both tests (arena test: mean bout length of trot; sudden object test: stand rate, next rate, and next percentage; Likert score for lead out when leaving the arena). Overall, these results indicate a limited influence of multiple testing, but a strong effect of preceding activity on the arena test. The comparison between solo and duo testing will require more research.

Introduction

The existence of personality in non-human animals is widely recognized (Gosling, 2008) and extensively studied in multiple species, even though there is no agreement on how to define and test it (David and Dall, 2016; Réale and Dingemanse, 2012). Personality research includes numerous studies on horses, as shown in the comprehensive review by König von Borstel (2013), which clearly demonstrates that existing personality tests and their interpretation are very diverse and that comparing different studies and applications is often complicated. Valid and reliable personality tests would certainly improve selection of horses for particular roles. More accurate selection would improve the safety of the people working with them and the welfare of the horses by ensuring that unsuitable animals were not subjected to challenging tests that they were unlikely to pass. Research on horse personality is a pivotal aspect of equitation science, which aims to study the behavior and performance of horses under human control and the effect of training and management on their welfare (McGreevy,

* Address for reprint requests and correspondence: Marc Pierard, KU Leuven Department of Biosystems, Leuven B 3001, Belgium. Tel.: +32 472571744.
E-mail address: marcalpierard@yahoo.com (M. Pierard).

https://doi.org/10.1016/j.jveb.2018.07.001
1558-7878/

2007). The further development of equitation science, as an applied scientific discipline, depends on the use of valid research methods and comprehensive reporting of methods and results (Pierard et al., 2015), especially if these results are to be translated into evidence-based applications.

Police horses face many different challenges in the workplace and so the specific selection and training of horses is crucial (Thomas, 2010). The mounted section of the Belgian federal police began to introduce a program of evidence-based changes to the management and training of their horses in 2011. From the start, optimizing the selection of highly suitable horses was a high priority and led to a research project to develop customized selection tests for potential new horses. These tests have been designed to complement existing veterinary examinations and subjective ratings by multiple experienced police riders during the standard two-week test period.

The first phase of this research project evaluated the feasibility of a customized test program and established an initial reference population of active police horses (Pierard et al., 2017). The second phase of the research project focused on the potential influence of the preceding activity on the results of the personality tests to establish if activity immediately before testing could confound tests results and therefore be a confounding variable that requires standardization. Horses were tested three times to compare three different preceding activities, which facilitated an analysis of the effect of retesting on each set of results. A pilot study was added to compare solitary testing to testing of two horses simultaneously to evaluate the effect of social isolation during standard testing.

Material and methods

Horses

The mounted section of the Belgian Federal Police is located in Etterbeek, a municipality of the Brussels-Capitol Region of Belgium. Of the approximately 140 Warmblood horses that were present at the time of the study, 30 were used in this study. The stable manager provided a list of available horses on test days. The person performing the tests did not know any of the horses before the start of the study. None of the horses had any experience with these personality tests before this study. The mean age of the test horses was 13.8 years (range 8–22) and their mean time at the mounted section was 9.8 years (range 3–19). The test group consisted of 11 mares and 19 geldings. Horses were housed in individual boxes.

The internal committee of the Belgian Federal Police, responsible for approving all equipment and procedures regarding welfare and safety for people and animals, approved this study and its protocol.

Test protocol

Tests were performed on 19 days between 22 June and 27 October 2013, always during weekends to minimize the interference with the regular operational duties of the mounted police. Testing took place in an indoor arena of 20 by 40 m with one entrance. The mirrors in the arena were covered to avoid any influence on the horses' behavior. During tests, the horses were videoed from a balcony above the entrance. The first 10 m after the entrance were fenced off with a barrier constructed from jump stands and poles to prevent the horses from traveling out of view of the camera during the tests, leaving a test area of 20 by 30 m. A blue and white umbrella was hanging from the ceiling in the middle of the arena and could be lowered and raised from the balcony remotely. Figure 1 shows a schematic overview of the test arena.

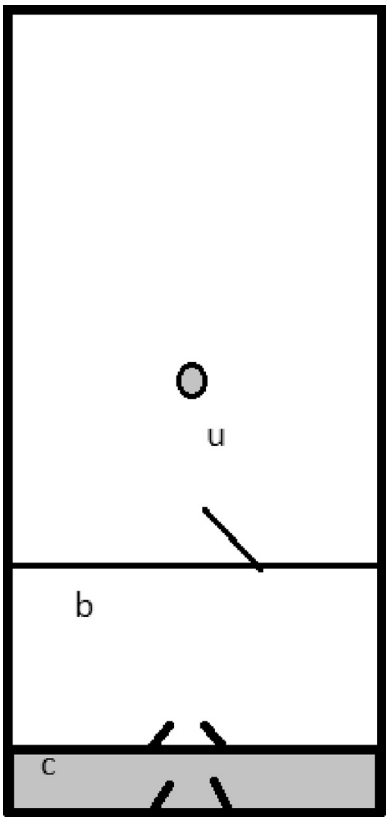

Figure 1. Floorplan of the test arena, showing positions of the umbrella (u), the barrier of jump stands and poles (b), and the position of the video camera (c) on the balcony above the entrance.

During tests, all horses were handled only by the first author, except for the duo tests, which required the assistance of a police officer to handle the second horse. The personality tests were an abbreviated version of the previously reported test protocol (Pierard et al., 2017). The order of testing and the variables measured are shown in Table 1. The variables were identical to the ones for the same test elements in Pierard et al. (2017). The test program started when the horse was led into the arena and through the opening in the barrier, after which it was cued to stop and stand still next to the handler for 1 minute. The horse was then released by taking the lead rope off the halter, followed by the handler closing the barrier and leaving the arena, closing the doors of the arena on the way out. As soon as the person reached the balcony, the horse would be observed for 10 minutes. At the end of this period the person would untie the rope from the balcony and lower

Table 1
The test program, including order of tests, variables scored in each test, and how they were measured

Test	Variables measured	Measuring method
Entering arena	Enter + stop/1 min stand still/ release	Likert score for each part
Arena	Stand/walk/trot/canter/roll/ whinny*/buck + jump*	Mean bout length, rate, percentage time (*only rate)
Sudden object	Stand/walk/trot/canter/roll/ whinny*/buck + jump*/ touching object/next to object/within 3m of object/ more than 3m from object	Mean bout length, rate, percentage time (*only rate)
Leaving arena	Approach + touch/attach lead rope/lead out	Likert score for each part

the umbrella until it reached the arena surface. The timing of the release of the umbrella was adapted to the position and activity of the horse to avoid hitting the horse with it. The umbrella was left on the ground for 5 minutes, and then raised again, with the horse being observed another minute afterward. Finally, the person would descend from the balcony, open the doors to the arena, enter the test zone, approach the horse, touch it on the shoulder, reattach the lead rope, and lead the horse out of the arena, opening the barrier in the process.

In total, 52 variables were measured (see Table 1), including six Likert scores, each ranging from 1 to 5, with 1 being the easiest and quietest and 5 indicating a horse that could not be controlled or approached. The entire test program was recorded with a digital video camera and videos were coded using BORIS 1.6 (Friard and Gamba, 2016). The ethogram applied in the analyses of the personality tests was drawn entirely from Pierard et al. (2017).

This experiment was set up as a crossover trial with three groups of 10 horses each. All horses were tested after three different preceding activities. These activities were: coming straight out of their individual box, being tied outside alone for about 30 minutes and spending about 30 minutes alone in a paddock of approximately 15 by 30 m. Each group was subjected to a different order of preceding activities, as shown in Table 2. Horses were tested only when they had been declared fit for duty, on days when they did not perform regular operational police work, and never twice on the same weekend. Latency between the three tests differed between horses, depending on availability, with only 10 of the 90 tests on consecutive weekends.

Most of the test horses were also videoed while tied (n = 21) or in the paddock (n = 27), with 18 horses being filmed during both preceding activities. Their behavior in the paddock was scored using the same ethogram as the arena test but while tied outside it was scored with a specific ethogram (see Table 3). For standing, moving, and pawing, measurements were mean bout length, rate and percentage of time, whereas whinnying was measured as only a rate.

At the end of the test period, 10 horses from the original sample were tested as five duos, immediately after emerging from their boxes. The composition of the duos was decided by the stable manager, selecting horses that were known to get along well to minimize risk of agonistic interactions and injuries.

Horses were familiar with the indoor arena and the paddock and were used to being tied outside for longer periods, during cleaning of their boxes, preparations for patrols, or while waiting to be deployed at various events.

Statistics

Data of the personality tests were analyzed with mixed models to evaluate treatment and order effects. Available data for behavior during the preceding activities (being tied and being in the paddock) were compared with behavior during the ensuing personality tests, using Spearman correlations. Behaviors while tied and in the paddock were also compared, again with Spearman correlations, for those horses filmed in both of these pretest activities. For horses used during the duo tests, behavior during the original personality tests straight out of the box was compared with behavior during the duo tests in two ways: by the Wilcoxon signed rank test and by Spearman correlations. Statistical analyses were performed with SAS 9.3 (SAS Institute Inc., Cary, North Carolina, USA, 2011).

Table 2
Order of preceding activities before personality tests for each group of horses

Group	Box	Tied	Paddock
Group A	1	2	3
Group B	3	1	2
Group C	2	3	1

Table 3
Ethogram for horses while tied

Behavior	Definition
Stand	Body not moving, three or four limbs in stance phase, or moving less than a full stride
Move	Horse moving at least one full stride
Paw	Protraction of one fore limb, followed by retraction, often scraping the floor during retraction, but also possible without contact with the floor
Whinny	"Loud, prolonged call, typically of 1 to 3 sec, beginning high pitched and ending lower pitched" (McDonnell, 2003)

Results

Order effects

Mixed models were used to evaluate the effect of test order and treatment. There was no interaction effect between order and treatment for any of the variables. Order effects tested for significant differences between the first, second, and third time the horses were tested, regardless of the preceding activity. Six of the 52 variables showed a significant order effect, distributed over the four parts of the tests. These were the Likert score for release ($P = 0.0002$) at entry, the rate of standing ($P = 0.003$) in the arena test, the rate of standing ($P = 0.002$) and walking ($P = 0.03$) and the mean bout length of walking ($P = 0.02$) in the sudden object test, and the Likert score for approach ($P = 0.0006$) when leaving the arena. The data for these variables and the comparisons per two tests are shown in Table 4.

Treatment effects

The treatment was the preceding activity (box, tied or paddock) and treatment effects tested for significant differences between the three different preceding activities, regardless of the order in which they were tested. This analysis revealed a significant difference in 10 of the 52 variables. These differences were found only for 10 variables for the arena test, that is, for none of the variables in the other parts of the test program. The behaviors with significant differences were standing (rate $P = 0.003$, mean bout length $P = 0.01$, percentage $P = 0.045$), cantering (rate $P = 0.0004$, mean bout length $P = 0.049$, percentage $P = 0.002$), rolling (rate $P < 0.0001$, mean bout length $P < 0.0001$, percentage $P < 0.0001$) and the bucking/jumping rate ($P = 0.0003$). The data for these variables and the comparison for each duo of treatments are shown in Table 5.

Correlations between preceding activity and tests

Data from observations before personality testing, while the horses were tied or in the paddock, were compared to behavior during the ensuing personality tests and to each other by using Spearman correlations. Table 6 shows the correlation coefficients ($P < 0.05$) for the behaviors with scores that correlated between the different parts of the personality tests, on the one hand, and while tied or in the paddock, on the other. Meanwhile, several variables while tied or in the paddock differed from some of the variables in parts of the personality tests. Table 7 contains significant correlation coefficients between such different variables. For the 18 horses

Table 4
Median and interquartile range (in brackets) of variables in the personality tests with significant order effects, with 2 × 2 comparison of test order (ns = not significant)

Test	Behavior	Test 1	Test 2	Test 3	Comparison (*P*)		
					Test 1-test 2	Test 1-test 3	Test 2-test 3
Arena	Stand rate	2.85/min (0.8)	2.4/min (0.7)	2.4/min (0.7)	0.0028	0.003	ns
Sudden object	Stand rate	2.55/min (0.9)	2.15/min (1)	2.25/min (0.8)	0.001	0.0029	ns
	Walk rate	3/min (1.3)	2.75/min (1.3)	2.6/min (1.4)	0.02	0.023	ns
	Walk bout	6.35s (4.3)	7.85s (5.5)	8.05s (4.8)	0.0054	0.048	ns
Entry	Release	1 (0)	1 (1)	2 (1)	ns	<0.0001	0.0031
Leave	Approach	2 (1)	1 (0)	1 (1)	0.003	0.0051	ns

that were filmed during both the time in the paddock and while tied, the significant correlation coefficients between variables for both activities are presented in Table 8.

Duo tests

The data for the 10 horses tested both alone (solo) and as duos were compared. The Wilcoxon signed rank test showed two significant differences ($P < 0.05$) between solo and duo testing: the mean bout length of standing in the arena test (solo: median = 8.25s interquartile range = 2.25; duo: median = 12.15s interquartile range = 15.83) and the score for entry/stop when entering the arena (solo median = 2 interquartile range = 1.025; duo: median = 1 interquartile range = 0.25).

Spearman correlations between the same variables during solo and duo testing, revealed five scores that correlated significantly in both conditions ($P < 0.05$): for mean bout length of trot ($r_s = 0.72$) in the arena test, during the sudden object test for stand rate ($r_s = 0.66$), next rate ($r_s = 0.84$) and next percentage ($r_s = 0.82$), and while leaving the arena for the score for lead out ($r_s = 0.75$).

Discussion

The main objective of this study was to evaluate if activity during the 30-min period before testing needed to be standardized as part of the overall protocol for the customized selection tests for police horses, described in Pierard et al. (2017). Because of practical limitations not all possible combinations of treatment orders were included in the crossover design and the original program of personality tests was abridged. The shortened test program did include handling (Likert scores at entering and leaving the arena), response to a sudden stimulus (sudden object test), and general activity (arena test), three major components of the full test protocol. Lansade et al. (2016) showed that the results of their full personality test and an abridged version, with an interval of one month, were significantly correlated for most variables. Because no horses from the previous phase were retested in our current experiment, no such relationship could be verified for our protocol.

The extraneous stimuli at the current site of testing, that was an urban location with associated activities, varied in their character and intensity. They included sounds from nearby traffic, local activity within and around the police installations, and the weather that probably affected horses most while tied or in the paddock. However, these horses were accustomed to these circumstances and all testing was conducted on weekend days, so the external influences were assumed to be randomly distributed and unlikely to have biased the current results. That said, controlling for and minimizing such stimuli would be preferred in future applications of the current approach.

Personality is often defined in terms of differences in individual behavioral responses that are consistent across time and situations (König von Borstel 2013; Réale and Dingemanse, 2012). This study evaluated consistency across different situations by altering the preceding activity, which differs from studying effects of variable situations, such as changing novel objects, during the actual tests themselves (Bulens et al., 2015; Christensen et al., 2008b, 2011). Because it has been shown that housing and breed can influence behavior during both arena and novel object tests (Lesimple et al., 2011), these factors were controlled for in our study by using only Warmblood horses that were all housed and managed the same way, at the same location.

Before actual testing, horses were never worked or exercised on test days, which was important because Werhahn et al. (2011) found that the behavior in a test paddock differed between horses that were trained before or after spending time in the paddock. Pessoa et al. (2016) showed that time in turnout had an effect on the behavior of police horses during and between patrols. Hausberger et al. (2011) reported that equestrian discipline affected behavior during arena and novel object tests, indicating an effect of training and management practices that differed between disciplines. They reported higher mean percentages of time standing in the arena test for sport horses (70.6%), compared to the current police horses (48.1%), but the results from their novel object test (54.8%) were very close to ours in the sudden object test (52.6%). They also found relatively low mean percentages of time spent cantering (arena test 4.4%; novel object test 3.8%), which are comparable to our results (arena test 4.7%; sudden object test 2.6 %).

Table 5
Median and interquartile range (in brackets) of variables in the arena test with significant treatment effects, with 2 × 2 comparison of treatments (ns = not significant)

Behavior		Preceding activity			Comparison (*P*)		
		Box	Paddock	Tied	Box-paddock	Box-tied	Paddock-tied
Stand	Rate	2.65/min (0.7)	2.4/min (0.5)	2.7/min (1)	0.0011	ns	0.0053
	Bout	9.6s (4.6)	13.8s (8.3)	9.4s (4.6)	0.016	ns	0.006
	%	48.2% (17.6)	57.7% (29.4)	42.45% (20.2)	ns	ns	0.012
Canter	Rate	0.3/min (0.5)	0/min (0.2)	0.4/min (0.8)	0.016	ns	<0.0001
	Bout	5.25s (4.1)	0s (3.8)	4.95s (5.4)	0.042	ns	0.026
	%	3.05% (6.8)	0% (2.1)	4.05% (10.8)	0.045	ns	0.0004
Roll	Rate	0.1/min (0)	0/min (0)	0.1/min (0)	<0.0001	ns	<0.0001
	Bout	11.55s (8.6)	0s (0)	11.65s (11.5)	<0.0001	ns	<0.0001
	%	2.2% (2.4)	0% (0)	2.15% (2)	<0.0001	ns	0.0001
Buck/Jump	Rate	0.4/min (0.7)	0/min (0)	0.35/min (0.8)	0.0012	ns	0.0001

Table 6
Significant Spearman correlations ($P < 0.05$) between variables for behavior in the paddock or while tied and the same behavior during the subsequent arena and sudden object test

Comparison test-activity	Variable	Spearman correlation
Arena/paddock	Whinny rate	0.63
Sudden object/paddock	Stand mean bout	0.54
	Stand percentage	0.4
	Walk rate	0.46
	Walk percentage	0.45
	Canter rate	0.39
Arena/tied	Whinny rate	0.46
Sudden object/tied	Whinny rate	0.45

Our results for treatment effects showed statistically significant differences only in the arena test. With 10 of the 17 variables affected, the situation in the 30 minutes immediately before the test program seems to have a particularly strong influence on the horse's behavior during the arena test. Compared to being in the box and being tied alone, horses behaved differently after time in the paddock: they stood less often but for longer and they also cantered, rolled, and bucked/jumped less. These results could indicate postinhibitory rebound in locomotory activity after the restricted conditions of box and tied. The presence of postinhibitory rebound in locomotory activity after confinement was reported by Freire et al. (2009) and Chaplin and Gretgrix (2010). Freire et al (2009) also reported that, in contrast, when compared with control situations, four forms of exercise (turnout, walker, riding, and treadmill) resulted in a significant decrease of locomotory activity during a rebound test, with less walking, trotting, cantering, rolling, and bucking (Freire et al., 2009), but testing was not immediately after the treatment but straight out of the box on days without exercise. Chaplin and Gretgrix (2010) found higher locomotory activity during the first five minutes after release in horses that had been boxed during the day and released into a paddock overnight, compared to horses that had been in a paddock full time. The results of these studies are compatible with our observations during the arena test. However, Freire et al. (2009) also mentioned that control horses were harder to handle for weighing than horses undergoing regular exercise. This is in contrast to our results, which revealed no significant differences in the ease with which horses could be handled at the start or the end of the test program. If postinhibitory rebound in locomotory activity was the only explanation for the differences in the arena test, we might expect it to also manifest as a drop in the ease of handling on entering the arena. If postinhibitory rebound in locomotory activity is contributing to the differences in the results of the arena test, it seems the effects lasts no longer than 10 minutes because no significant treatment differences were present in the subsequent sudden object test.

Our analyses did not reveal any significant differences between tests straight out of the box and after being tied outside. This could indicate that both treatments are experienced in a similar way by the horses but this cannot be confirmed until additional research has explored different forms of confinement and their effect on behavior.

The crossover design also allowed us to evaluate whether repeating the tests multiple times affected test results and whether the personality tests showed consistency over time. Multiple testing over a period of up to 4 months seemed to have a limited influence, as 46 of the 52 variables showed no significant differences. For the arena and sudden object tests, there were no significant differences between Test 2 and Test 3. When tested the first time, the horses were observed standing more often than the second and third time in both arena and sudden object tests and, in the sudden object test, they walked more often but in shorter bouts. For the entry and leave observations, horses showed a more active response on being released in the third test, than in the first two, and were more easily approached in Test 2 and 3, than in Test 1, which could have been affected by increased familiarity with the handler. Seaman et al. (2002) repeated several personality tests three times with an average of 9 days between retesting and found that the arena test was more consistent than tests involving a person or a novel object. This last conclusion aligns with our results, in which only the sudden object test had more than one variable with significant differences. However, Seaman et al. (2002) reported increased latency times to approach a person or a novel object in their third test, compared to their first or second tests, partly caused by more animals not approaching at all within the time limits. In our study, we did not include latency times but we did not find any significant differences in variables for distance categories in the sudden object test.

Multiple testing has also been studied for young horses, which are still maturing physically. Lansade et al. (2008a) reported

Table 7
Significant Spearman correlations ($P < 0.05$) between different variables for behavior while tied or in the paddock, and during subsequent personality tests

Compared test/activity	Variable test	Variable activity	Spearman correlation
Arena/tied	Stand mean bout	Move rate	−0.46
	Walk percentage	Paw mean bout	−0.45
	Roll rate	Move percentage	−0.47
	Roll percentage	Stand percentage	0.52
		Move rate	−0.42
		Move mean bout	−0.47
		Move percentage	−0.56
	Roll rate	Move percentage	−0.47
	Roll mean bout	Paw mean bout	−0.47
	Canter rate	Whinny rate	0.47
	Canter percentage		0.46
Sudden object/tied	Walk rate	Move rate	0.45
	Walk rate	Move percentage	0.49
	Walk percentage		0.44
	Trot bout	Paw mean bout	0.47
	Trot percentage		0.49
Entry scores/tied	1 min stand still	Move mean bout	0.46
		Move percentage	0.45
		Whinny	−0.46
	Release	Stand percentage	−0.51
		Move percentage	0.53
Entry scores/paddock	Release	Stand rate	0.46
		Stand mean bout	−0.51
		Stand percentage	−0.42
		Walk rate	0.56
		Walk percentage	0.51
Leaving arena scores/paddock	Attach lead rope	Trot percentage	0.48

Table 8
Significant Spearman correlations ($P < 0.05$) between variables for behavior in the paddock and while tied

Variable paddock	Variable tied	Spearman correlation
Trot rate	Move mean bout	−0.56
Trot percentage		−0.54
Canter rate		−0.61
Canter mean bout		−0.69
Canter percentage		−0.68
Buck/jump rate	Move mean bout	−0.78
Roll rate	Paw rate	0.48
Roll mean bout	Paw mean bout	−0.58
	Paw percentage	−0.48
Whinny rate	Whinny rate	0.53

consistent reactions to fear-inducing situations among young horses between 8 months and 2.5 years, with intervals between testing of 10 and 12 months. For the same horses, they also reported consistency over time for responses to social isolation (Lansade et al., 2008b), with the rate of whinnying being the most reliable variable. This too is in accordance with our results, which show no treatment or order effect for whinnying but significant correlations for most comparisons. Visser et al. (2001) found variable consistency over time in young Warmblood horses, tested on four occasions (at age 9, 10, 21, 22 months), with lower consistency when comparing the first to the second year.

Most of the test horses were videoed while in the paddock or tied outside, allowing comparisons of their behavior in these contexts to their subsequent behavior during the personality tests. As mentioned, the rate of whinnying was significantly positively correlated between each combination of tests in which it was measured, suggesting this could be a variable that is consistent across situations. Whinnying also showed significant correlations between different elements of the full testing protocol in the previous phase of this research project (Pierard et al., 2017). Our current finding that whinnying was the only significantly correlated behavior between the arena test and the paddock treatment supports the conclusion that the arena test is most influenced by preceding activity. Comparing tests to the tied treatment is more complicated due to it permitting different variables. The more horses moved while tied, the less they would roll and the more they would walk during the arena test. Given that there were no significant differences between horses tested straight out of the box and after being tied, it would be interesting to explore possible relationships between behavior in their box and while tied. In this study, it was not possible to video record the behavior of horses in their box.

For the entry test, there was no significant correlation of behavior while tied with the score for leading the horse in through the barriers and stopping, but the horses that were easier to handle during 1 minute of immobility and upon release, moved less while tied, suggesting that responses to restrictions of mobility are related in different situations. Comparing the entry test to behavior while in the paddock revealed that horses that responded calmer on release in the arena were, in the paddock, standing less often but for longer and walking less often but longer in total, which means they spent a greater proportion of time in the paddock showing the least active behaviors. Finally, comparing behavior in the paddock to that while tied showed the unexpected result that longer mean bout duration of moving while tied was linked to less trotting, cantering, and bucking/jumping in the paddock.

After prior testing using the full protocol (Pierard et al., 2017), some riders were surprised at the reactions of the horse with which they mostly worked, especially in the arena and sudden object test. Riders suggested the solitary testing could have made a critical difference to their behavior because the horses were almost always worked and trained with at least one other horse. Accordingly, in this study, we tested compatible horses simultaneously. The results can be considered a test of consistency across contexts because the presence of conspecifics is part of the context (Stamps and Groothuis, 2010). When comparing the 10 horses in the solo and duo test, only two of the 52 variables differed significantly, indicating that 50 variables were not significantly influenced by the presence of a conspecific. One of the variables that showed a significant difference was the score for entering the arena and stopping. Horses were somewhat harder to lead and stop when alone than when they had a companion, possibly indicating that company can reduce fear or arousal. However, when calculating Spearman correlations between solo and duo testing, only five variables correlated significantly, whereas 47 did not.

Christensen et al. (2008a) reported that test horses reacted less to a suddenly moving object when in the company of a previously habituated horse than when in the company of a naïve horse. This was confirmed with lowered fear responses for foals that were exposed to frightening stimuli in the company of their previously habituated mother, as compared to foals who did not get such exposure (Christensen, 2016). As these police horses had received extensive training to habituate them to a wide range of potentially scary objects, they could be acting as a calm and habituated companion. The effect of testing conspecifics simultaneously will need more study, including the potential role of the relation between the individuals included or at least their familiarity with one another.

Conclusions

The results of this experiment indicate that testing horses multiple times with the tests described is feasible without having a major influence on the result of later tests, but that activity before testing merits consideration, especially for the arena test as it occurs in the current test program. Pretest activity, especially in the 30 minutes immediately before testing, will have to be standardized as much as possible to produce reliable personality profiles for horses being tested as potential police horses. The effect of turnout in paddocks on subsequent behavior is also relevant for the welfare of horses and the safety of police riders.

Acknowledgment

The authors wish to thank the mounted section of the Belgian Federal Police for their continued support and cooperation by allowing the use of their horses and infrastructure. The help of a police officer of the mounted police during the duo testing was greatly appreciated. Professor Geert Verbeke advised on statistical analysis of the data.

This article was written by all authors, who have approved the final version. The tests were designed and performed by Marc Pierard. All authors contributed to the analysis and interpretation of the data.

Conflict of interest

None of the authors declared a conflict of interest.

References

Bulens, A., Sterken, H., Van Beirendonck, S., Van Thielen, J., Driessen, B., 2015. The use of different objects during a novel object test in stabled horses. J. Vet. Behav.: Clin. Appl. Res. 10, 54–58.

Chaplin, S.J., Gretgrix, L., 2010. Effect of housing conditions on activity and lying behaviour of horses. Animal 4, 792–795.

Christensen, J.W., 2016. Early-life object exposure with a habituated mother reduces fear reactions in foals. Anim. Cogn. 19, 171–179.

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

Christensen, J.W., Zharkikh, T., Ladewig, J., 2008b. Do horses generalise between objects during habituation? Appl. Anim. Behav. Sci. 114, 509–520.

Christensen, J.W., Zharkikh, T., Chovaux, E., 2011. Object recognition and generalisation during habituation in horses. Appl. Anim. Behav. Sci. 129, 83–91.

David, M., Dall, S.R.X., 2016. Unravelling the philosophies underlying "animal personality" studies: a brief re-appraisal of the field. Ethology 122, 1–9.

Freire, R., Buckley, P., Cooper, J.J., 2009. Effects of different forms of exercise on post inhibitory rebound and unwanted behaviour in stabled horses. Equine Vet. J. 41, 487–492.

Friard, O., Gamba, M., 2016. BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. Methods Ecol. Evol. 7, 1325–1330.

Gosling, S.D., 2008. Personality in non-human animals. Soc. Personal. Psychol. Compass 2, 985–1001.

Hausberger, M., Muller, C., Lunel, C., 2011. Does work affect personality? A study in horses. PLoS One 6, e14659.

König von Borstel, U., 2013. Assessing and influencing personality for improvement of animal welfare: a review of equine studies. CAB Rev. 8, 1–27.
Lansade, L., Bouissou, M.F., Erhard, H.W., 2008a. Fearfulness in horses: a temperament trait stable across time and situations. Appl. Anim. Behav. Sci. 115, 182–200.
Lansade, L., Bouissou, M.-F., Erhard, H.W., 2008b. Reactivity to isolation and association with conspecifics: a temperament trait stable across time and situations. Appl. Anim. Behav. Sci. 109, 355–373.
Lansade, L., Philippon, P., Hervé, L., Vidament, M., 2016. Development of personality tests to use in the field, stable over time and across situations, and linked to horses' show jumping performance. Appl. Anim. Behav. Sci. 176, 43–51.
Lesimple, C., Fureix, C., Lescolan, N., Richard-Yris, M.-A., Hausberger, M., 2011. Housing conditions and breed are associated with emotionality and cognitive abilities in riding school horses. Appl. Anim. Behav. Sci. 129, 92–99.
McDonnell, S.M., 2003. The Equid Ethogram: A Practical Field Guide to Horse Behavior. Eclipse Press, Lexington, KY.
McGreevy, P.D., 2007. The advent of equitation science. Vet. J. 174, 492–500.
Pessoa, G.O., Trigo, P., Mesquita Neto, F.D., Lacreta Junior, A.C.C., Sousa, T.M., Muniz, J.A., Moura, R.S., 2016. Comparative well-being of horses kept under total or partial confinement prior to employment for mounted patrols. Appl. Anim. Behav. Sci. 184, 51–58.
Pierard, M., Hall, C., König von Borstel, U., Averis, A., Hawson, L., McLean, A., Nevison, C., Visser, K., McGreevy, P., 2015. Evolving protocols for research in equitation science. J. Vet. Behav.: Clin. Appl. Res. 10, 255–266.
Pierard, M., McGreevy, P., Geers, R., 2017. Developing behavioral tests to support selection of police horses. J. Vet. Behav.: Clin. Appl. Res. 19, 7–13.
Réale, D., Dingemanse, N.J., 2012. Animal Personality. In: ELS. John Wiley Sons, Ltd, Chichester, pp. 1–8. Available at: https://doi.org/10.1002/9780470015902.a0023570.
Seaman, S.C., Davidson, H.P.B., Waran, N.K., 2002. How reliable is temperament assessment in the domestic horse (*Equus caballus*)? Appl. Anim. Behav. Sci. 78, 175–191.
Stamps, J., Groothuis, T.G.G.G., 2010. The development of animal personality: relevance, concepts and perspectives. Biol. Rev. 85, 301–325.
Thomas, R., 2010. Predictability in an unpredictable environment: training the police horse using learning theory. J. Vet. Behav.: Clin. Appl. Res. 5, 218.
Visser, E.K., Van Reenen, C.G., Hopster, H., Schilder, M.B.H., Knaap, J.H., Barneveld, A., Blokhuis, H.J., 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. Appl. Anim. Behav. Sci. 74, 241–258.
Werhahn, H., Hessel, E.F., Schulze, H., Van den Weghe, H.F.A., 2011. Temporary turnout for free exercise in groups: effects on the behavior of competition horses housed in single stalls. J. Equine Vet. Sci. 31, 417–425.