

Automated real-time stress monitoring of police horses using wearable technology

by Norton, T., Piette, D., Exadaktlyos, V. and Berckmans, D.

Copyright, Publisher and Additional Information: This is the author accepted manuscript. The final published version (version of record) is available online via Elsevier. Please refer to any applicable terms of use of the publisher.

DOI: <https://doi.org/10.1016/j.applanim.2017.09.009>



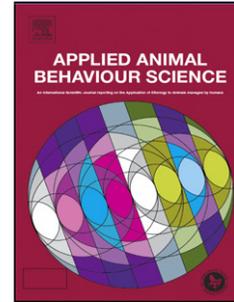
Norton, T., Piette, D., Exadaktlyos, V. and Berckmans, D. 2017. Automated real-time stress monitoring of police horses using wearable technology. *Applied Animal Behaviour Science*

13 October 2017

Accepted Manuscript

Title: Automated real-time stress monitoring of police horses using wearable technology

Authors: Tomas Norton, Deborah Piette, Vasileios Exadaktylos, Daniel Berckmans



PII: S0168-1591(17)30259-9
DOI: <https://doi.org/10.1016/j.applanim.2017.09.009>
Reference: APPLAN 4515

To appear in: *APPLAN*

Received date: 9-1-2017
Revised date: 7-9-2017
Accepted date: 17-9-2017

Please cite this article as: Norton, Tomas, Piette, Deborah, Exadaktylos, Vasileios, Berckmans, Daniel, Automated real-time stress monitoring of police horses using wearable technology. *Applied Animal Behaviour Science* <https://doi.org/10.1016/j.applanim.2017.09.009>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Automated real-time stress monitoring of police horses using wearable technology

Tomas Norton^{1,2}, Deborah Piette¹, Vasileios Exadaktylos³, Daniel Berckmans¹

¹ M3-BIORES, Division Animal and Human Health Engineering, Department of Biosystems,
KU Leuven University, Oude Markt 13 3000 Leuven, Belgium

² Agricultural Engineering Innovation Centre, Department of Engineering, Harper Adams
University, Newport, Shropshire, TF10 8NB, United Kingdom

³ BioRICS NV, Technologielaan 3 3001 Heverlee, Belgium

Corresponding author:

Daniel Berckmans

Email: daniel.berckmans@kuleuven.be

Full postal address: Kasteelpark Arenberg 30, 3001 Heverlee, Belgium

Telephone: +32 16 32 17 26

Person to address during the submission and review process:

Deborah Piette - deborah.piette@kuleuven.be

Full postal address: Kasteelpark Arenberg 30, 3001 Heverlee, Belgium

Telephone: +32 16 37 46 67

Highlights

- Subjective evaluation of police horses differed most between police riders when new horses were evaluated at the mounted police in Brussels, Belgium.
- A scientifically based tool to evaluate future performance of candidate police horses could help the mounted police to improve their selection procedure for police horses, which is mainly subjective.
- A methodology for automated stress monitoring in horses using wearable technology is presented in this paper. The method performs with a sensitivity of 77% and a precision of 78%.
- For one of the four protocols, the time spent in a relative stress zone of 20% was significantly lower for good beginner horses compared to bad beginner horses
- This study demonstrated that real-time stress detection with wearable technology in mounted police horses provides information on the longer term suitability of police horses

1 Abstract

Mounted police horses and riders are repeatedly subjected to demanding and stressful situations. Intensive selection and training of police horses is required to ensure performance, safety and welfare of the horses and their riders. At the mounted police in Brussels, Belgium, the selection of police horses is mainly based on intuition built upon previous experience which makes it a subjective decision. Although this decision mostly leads to good results, sometimes horses are purchased that turn out to be unsuited for the mounted police. Including an objective measure when evaluating the longer term suitability of candidate police horses could help the mounted police to further improve their selection procedure for police horses. This study

investigated the potential of real-time stress detection as a measure to evaluate the longer term suitability of police horses. Four experimental protocols were developed in consultation with the mounted police and were performed by 17 horse-rider pairs. The horses were divided into four categories according to their experience and suitability as police horses: good beginner, bad beginner, good experienced and bad experienced. The relative stress of the horses was monitored with wearable technology during every protocol. For one protocol, the time percentage spent over 20% relative stress by the horse was found to be significantly lower for good compared to bad beginner horses (p-value = 0.0277). In conclusion this study demonstrated that real-time stress detection with wearable technology in mounted police horses provides information on the longer term suitability of police horses.

Keywords: horse, police, stress, performance, suitability, wearable technology

2 Introduction

Horses are still used in military and police functions. Although most military horses are only used for ceremonial purposes, many countries still have an operational mounted police unit (Rawdon-Mogg, 2016). The main tasks of these mounted police units are to assist in interventions and to maintain order in public places (Federale Politie, 2014; Police Nationale, 2017; Royal Newfoundland Constabulary, 2017). When police horses are on duty, they have to cope with stressful situations. Intensive training is required to teach them how to remain calm at all times to ensure their own safety as well as that of their riders. However, for training to be effective the horses themselves must be properly selected. This is important because the way a horse reacts to stressful events does not solely depend on training, but is also related to their breed, temperament and earlier experiences in life (von Borstel, et al., 2011; Munsters, et al., 2013; Sankey, et al., 2010; Valenchon, et al., 2013).

At the mounted police in Brussels, the selection of new police horses is primarily based on the intuition built upon previous experience of the mounted police riders and trainers. The high subjectivity of this decision does not, however, imply that it is a bad decision. In fact, at the mounted police in Brussels, selected horses usually live up to their expectations. Nevertheless, occasionally a horse is bought that does not turn into a good police horse, which means loss of a high investment and a lot of time. This is where an additional tool to assess the suitability of candidate police horses might prove useful.

Several studies have tried to facilitate the selection procedure for police horses (Flentje, et al., 2010; Pierard, et al., 2014; Pierard, et al., 2013; Thomas, 2010). The objective of most of these studies is to predict future performance and evaluate the suitability of police horses in a more objective manner. Although considerable research has already been done, there are still a few gaps in scientific literature. The first one being that the horses are not always tested in contexts that represent police situations, which could lead to the questioning of the relevance of these experiments. The second gap is that a method to objectively predict the suitability of a horse as a police horse still has not been established. For instance, Flentje (2010) tried to find reliable temperament tests to predict the suitability of a horse as a police horse (Flentje, et al., 2010). The study tested horses' reliability of response to several repetitions of a social isolation test, a novel stimulus test and a handling test. The findings showed that these temperament tests do not succeed in predicting the suitability of horses as police horses due to lack of reliability.

Another study was performed by Pierard (2013, 2014 & 2017), who worked out a program for the mounted police in Brussels, consisting of temperament tests in which the horse is running free such as a sudden noise test. By evaluating measures like time spent cantering in relation to the distance to the stimulus, Pierard worked out a "framework to evaluate new horses" (Pierard, et al., 2013). In his study, Pierard demonstrated the feasibility of a custom test protocol, but did not find extensive differences between well-suited and less-suited police

horses. Although these temperament tests do provide the mounted police with a means for more objective selection of candidate police horses, the test situations used by Pierard differ from actual mounted police situations. Mounted police horses never run free when they are on duty. Since the presence of a rider can influence the behaviour of the horse, testing candidate police horses in ridden situations might be more relevant (von Borstel, et al., 2011).

This concept of testing horses in relevant situations was suggested in research performed by von Borstel (2011), in which equine heart rate and behaviour were observed during temperament tests. These tests were conducted with the horse being ridden, being hand-led and running free. A poor correlation was found between the measured variables for free-running and riding tests. According to the authors “this might indicate that it may be more important to measure the horse during those situations that are most relevant in practice, i.e. riding” (von Borstel, et al., 2011).

Munsters (2013) presents what seems to be the only research study in which police horses are evaluated in ridden situations (Munsters, et al., 2013). Ridden police horses with different levels of experience were subjected to a series of tests. The heart rate and heart rate variability of the horses was measured, their behaviour was scored, and a police performance score was given to every horse by two experienced mounted police trainers. One objective of the study was to examine the correspondence between the rather subjective behaviour and police performance scores and the more objective heart rate measures. A second goal of the study was to see if experienced and inexperienced police horses showed different behavioural and physiological responses. The study found that police performance score showed more correspondence with heart rate and heart rate variability, compared to behavioural score. However, no significant differences could be detected between the inexperienced and experienced police horses using any of the physiological or behavioural measures.

In summary we can conclude from literature that to date no objective standard procedure is available that successfully predicts whether or not a horse is suitable for the mounted police. The hypothesis of this study is that the stress level of a horse can serve as indicator of its long term suitability as a police horse. Experiments were conducted with horses of different categories reflecting different levels of experience and suitability. The relative stress of horses during the experiments was monitored in real-time using wearable technology that measures heart rate and activity. Based on the horses' relative stress, several features were identified that can provide information on the longer term suitability of candidate police horses. The first objective of this work was to develop experimental protocols that are representative of situations at the mounted police in Brussels. The second objective of this study was to search for an objective measure that can be used to evaluate the longer term suitability of horses as a police horse.

3 Materials and methods

3.1 Protocols

Four protocols were defined in agreement with the mounted police. All four protocols represent an aspect of the everyday mounted police trainings and practices, so that the experiments are representative and convenient for the mounted police. The 'going to work' test (GTW) serves to observe the reaction of the horse when it is prepared to go to work and when it is brought to the training arena (Table 1). The flexibility test (FT) is designed to examine the reaction of the horse to a change in rider (Table 2). The goal of the obstacle test (OT) is to observe the horse's reaction to a course of obstacles (Table 3, Figure 1). The 'coming from work' test (CFW) serves to observe the horse's reaction when it comes back from the training arena and is prepared to go back into its box (Table 4).

The four protocols were performed consecutively in the following order by a certain horse-rider pair: GTW, OT, FT, CFW. In this way, a workflow was created that closely matches a normal training session at the mounted police. Between protocols, horse and rider paused for only a few minutes so that the connection of the sensors could be checked by the researcher. On average a little over one hour passed between the start and the end of each horse-rider pair measurement. During every experiment an additional rider or trainer from the mounted police, further referred to as ‘instructor’, was present in the arena as the second rider for the FT. The four protocols were approved by the Animal Ethics Committee of KU Leuven.

Table 1: Protocol ‘Going To Work’ (GTW)

Step	Description
1	The horse is led to the cleaning area, the sensors are attached to the horse according to the instructions given in the paragraph ‘Equipment’
2	The horse is left alone during 3 minutes to measure its resting heart rate
3	The horse is prepared to go to work: brushing, saddling, etc.
4	The rider leads the horse to the arena and mounts
5	The horse stands still for 20 seconds
6	The horse walks around in the arena at a steady pace during 3 minutes

Table 2: Protocol ‘Flexibility Test’ (FT)

Step	Description
1	The horse (already equipped with all sensors) stands still for 20 seconds
2	The horse walks around in the arena at a steady pace during 4 minutes
3	The rider leads the horse to the middle of the arena, stops and dismounts
4	The second rider enters the arena and mounts
5	The first rider leaves the arena
6	The horse walks around in the arena at a steady pace during 4 minutes

Table 3: Protocol ‘Obstacle Test’ (OT)

Step	Description
1	The horse (already equipped with all sensors) stands still for 20 seconds
2	The horse walks around in the arena at a steady pace during 3 minutes

3	<p>The horse completes the obstacle course on the centreline of the arena while walking, if the horse stops or starts trotting or cantering, it is asked by the rider to walk again. The obstacle course is illustrated in Figure 1 and consists of:</p> <ul style="list-style-type: none">– Walking over plastic bottles– Walking through a half open curtain of plastic straps– Knocking over a barrel (the rider knocks the barrel over, not the horse) <i>Four blue barrels are stacked in pairs of two alongside the obstacle course, two on the left side of the horse and rider, two on the right side (Figure 1). The rider knocks over the upper barrel on the right hand side.</i>– Stopping by the speaker while a sound is played <i>The horse stops with its front legs at the level of speaker. The instructor in the arena then presses the button to play the sound. The sound recording lasts 20 seconds and is a crescendo consisting of applause, whistles, gun sounds and explosions.</i>
4	The horse walks around in the arena at a steady pace during 3 minutes

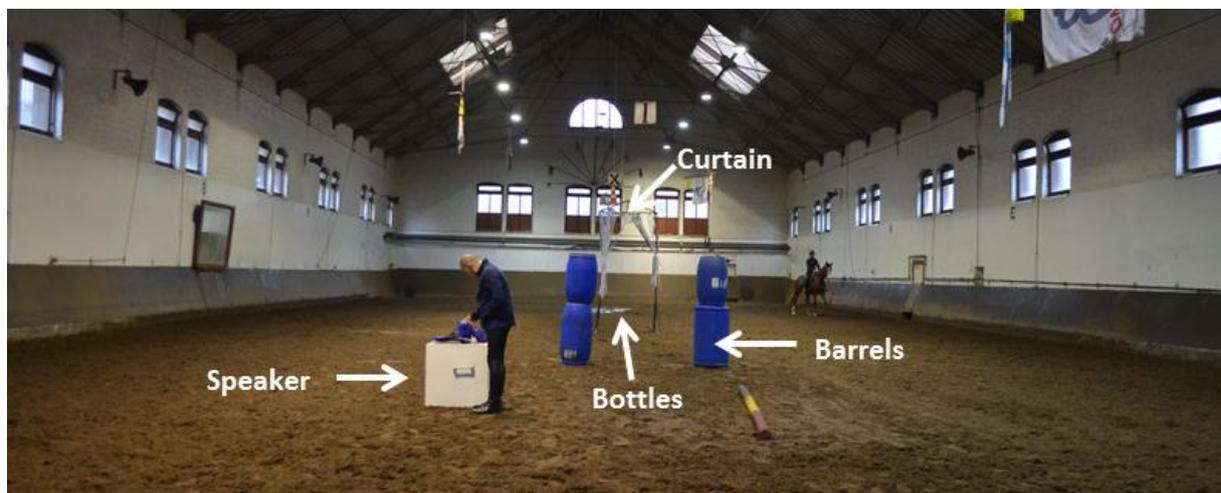


Figure 1: The obstacle course

Table 4: Protocol ‘Coming From Work’ (CFW)

Step	Description
1	The horse (already equipped with all sensors) stands still for 20 seconds
2	The horse walks around in the arena at a steady pace during 3 minutes
3	The rider stops the horse, dismounts and leads the horse the cleaning area
4	The horse is attached in the cleaning area
5	The horse is left alone for 3 minutes
6	The horse is prepared to go back to its box: the saddle is taken off, etc.
7	The horse is left alone for 3 minutes
8	The sensors are removed from the horse according to the previously given instructions and the horse is put in its box

3.2 Subjects and infrastructure

Detailed information on the horses can be found in Table 5. The horses participating in the measurements were subdivided into four categories depending on their experience and suitability as police horse: Good Beginner (GB), Bad Beginner (BB), Good Experienced (GE) and Bad Experienced (BE). Beginner horses are horses that have recently (20 days– 4 months) entered the training period at the mounted police. Experienced horses, in contrast, have been operational for 1 year up to 10.5 years. The label ‘good’ or ‘bad’ has been given to the horses previous to the start of the experiments by the mounted police riders based on their subjective experience with the horses and based on the progress the horses have been making during their training period or operational career. Within each category two to three horses performed the four protocols twice. In order to take into account the impact of the rider on the horse, the repetition of the four protocols with a certain horse was carried out with a different rider than the one in the first measurement. For the category ‘GB’ the repetition with one of the three horses could not be completed because the horse got injured during the experiments. The horses were ridden by nine different mounted police riders with different years of experience in horse riding. Information about the riders is listed in Table 6. An overview of the combinations of riders and horses during the measurements can be found in Table I of the supplementary material. In total 17 horse-rider pairs participated in the study.

The experiments were performed in an indoor arena (21 m x 55 m) with sand and fibre bedding. During the experiments the horses were kept on straw in indoor boxes (3.5 m x 2.95 m). They were fed a normal diet of four kilos of grain mix, spread over three times during the day and received five kilos of hay twice a day. The working hours of the horses varied between one to twelve hours a day (transportation included) depending on whether or not they were on duty. The day that a horse participated in the experiment, it was not on duty.

Table 5: Detailed information about the horses

Horse ^a	Repetitions	Sex	Age (years)	Breed	Colour	Height (cm)	Weight ^b (kg)
GB-1	2	gelding	9	unknown	chestnut	169.5	600
GB-2	1	mare	4	Drum horse	piebald	167	579
GB-3	2	gelding	4	BWP ^c	black	164	496
BB-1	2	gelding	5	Zangersheide	brown	170	601
BB-2	2	mare	4	sBs ^d	grey	161.5	540
GE-1	2	gelding	6	BWP	grey	173	586
GE-2	2	gelding	9	BWP	brown	174.5	562
BE-1	2	mare	13	KWPN ^e	chestnut	169.5	546
BE-2	2	gelding	15	sBs	chestnut	173	575

^a GB: Good Beginner – BB: Bad Beginner – GE: Good Experienced – BE: Bad Experienced

^b Approximate weight in kg: (body length (cm) x girth (cm) x girth (cm)) / 11877 (Heipertz-Hengst 2002)

^c Belgian Warmblood

^d Belgian sport horse

^e Dutch Warmblood

Table 6: Detailed information about the riders

Rider	Gender	Age / Horse riding experience (years)	Rider	Gender	Age / Horse riding experience (years)
R-1	female	29 / 12	R-6	female	24 / 3
R-2	female	23 / 11	R-7	female	26 / 3
R-3	male	38 / 4	R-8	male	45 / 26
R-4	male	50 / 30	R-9	male	35 / 15
R-5	male	27 / 5			

3.3 Measuring equipment and software

The heart rate of the horses was measured using the Polar Equine RS800CX Science set from Polar Electro Belgium. The activity of the horses was measured using the Zephyr BioHarness™ 2.0 module (Zephyr Technology Corp.) which was attached at the bottom of the horse's belly using a custom made strap with pocket. Heart rate and activity were both measured at a sampling rate of 1 Hertz in beats per minute (bpm) and Vector Magnitude Units (VMU) respectively. The equipment was attached to the horse by the researcher. All experiments were recorded on video using a Sony Handycam Camcorder. These recordings served to analyse the horses' behaviour in detail at the end of the measurements using an equine ethogram. More specifically, a list of thirteen types of negative equine behaviour chosen from the most frequently used ones in literature has been composed. The types of behaviour that were included in the ethogram are: ears focussed (ears clearly pinned towards an object or sound), ears flat (ears against the neck), tail beating (sudden and fierce up and down or sideways movements), tail high (tail higher than the croup), head high (eyes higher than the withers), head tossing (sudden and fierce up and down movements), bucking (both hind legs kicking), rearing (standing on hind legs), hitting (kicking with one leg), jumping (sudden jump forward or sideways), refusing (clearly not obeying the rider's signals), fleeing (running away from object or sound) and scraping (scraping with front leg). It was chosen to evaluate only negative behaviour since the negative stress of the horses is of particular interest in this study. Using the video recordings of the horses together with a labelling tool developed by Viazzi (Viazzi, et al., 2011), the occurrence of each type of behaviour was stored for every second of the experiment. This resulted in equine behaviour data with the same 'sampling rate' as the measured heart rate and activity, namely 1Hz. The amount of different types of behaviour were summed per second in order to obtain one negative behaviour graph per measurement. The software used for data analysis included Matlab 2014a (Matlab) and the Captain Toolbox (Taylor, et al., 2007).

3.4 Evaluation of the horse by the rider and the instructor

In addition to the evaluation of the horse's behaviour using an equine ethogram, the instructor present in the arena during the experiment and the rider were asked to evaluate the horse as well. The evaluation form comprises four statements that must be scored on a scale from zero to ten; zero meaning not agreeing with the statement at all, and ten meaning fully agreeing with the statement. The statements were the following: 1) The horse was willing, 2) The horse was obedient, 3) The horse was calm, 4) The horse was nervous. Statement 1 to 3 are positive statements, statement 4 is a negative statement. The evaluation was completed by the rider and the instructor after the horse-rider pair had gone through all four protocols.

3.5 Stress detection

The heart rate of a living organism can be decoupled into three major components. A first component is related to the basal metabolism of the organism, a second component correlates with the organism's physical activity and a third component links to the organism's stress. As already mentioned the scope of this study is to monitor the stress of horses. In order to extract this stress component in the heart rate, the components 'physical activity' and 'basic metabolism' must be quantified. In this work, the physical heart rate component of the horses and their riders is estimated based on their measured activity. Their basic metabolism is assumed constant during the short experiments (8 to 25 minutes) and is consequently not further taken into account.

Using the measured physical activity, a modelled heart rate can be obtained that only reflects changes in heart rate resulting from changes in physical activity (Jansen, et al., 2009). This is done by regarding the horse as a Single-Input-Single-Output (SISO) system with activity as input variable and heart rate as output variable (Jansen, et al., 2009). Consequently the relation

between input and output variable can be described by a linear AutoRegressive model with eXogenous input (ARX) of which the general equation is given by:

$$y(t) + a_1y(t - 1) + \dots + a_ny(t - n) = b_0u(t) + b_1u(t - 1) + \dots + b_my(t - m)$$

with:

y : output variable (measured heart rate)

u : input variable (measured activity)

t : time instant of the measurement

a_1, \dots, a_n : a-parameters

b_0, b_1, \dots, b_m : b-parameters

For this work, an ARX model of the second order with a time delay of one second was used. This resulted in a modelled heart rate describing the horse's heart rate component that is solely related to activity. This is in contrast to the measured heart rate which is also dependent on stress. The positive difference between the measured total heart rate and the modelled activity related heart rate is a measure for the horse's stress. After applying a 20 second window averaging filter to this difference and normalising the result, a relative stress graph is obtained for every protocol. The relative stress graph shows the stress of the horse in percentage, indicating 0% when the horse is not stressed at all and 100% when the horse is maximally stressed during the protocol. The different steps of stress detection are illustrated in Figure 2.

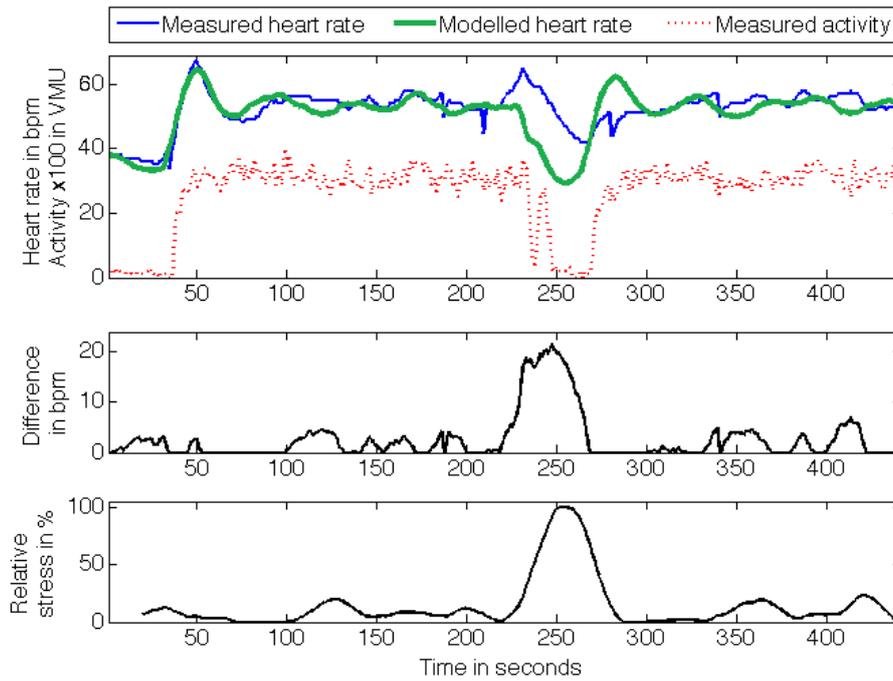


Figure 2: Illustration of the stress detection. From top to bottom: modelled heart rate calculated based on physical activity, compared with measured heart rate – positive difference between measured and modelled heart rate – relative stress in % after applying window averaging and normalisation.

The same window averaging and normalisation was applied to the negative behaviour graph of the horses in order to obtain a smoother relative negative behaviour graph in percentage. This graph shows 0% when the horse is not showing any negative behaviour and 100% when the horse is showing maximal negative behaviour during the protocol, as illustrated in Figure 3.

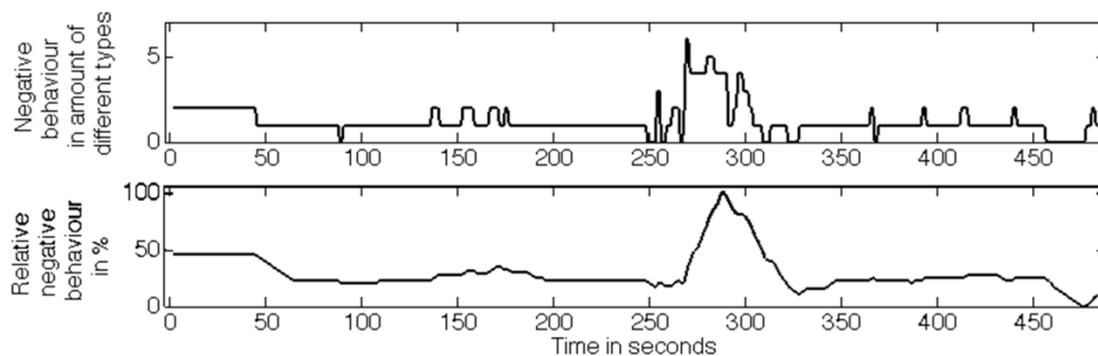


Figure 3: Obtain the relative negative behaviour in percentage from the labelled negative behaviour

For the validation of the method for stress detection, the relative stress was compared with the relative negative behaviour. A peak detection algorithm was applied to the computed stress graph and the behaviour graph (minimal peak distance = 50 seconds, minimal peak height = 10%) and positive peaks were matched between both time series in a range of ± 50 seconds.

Peaks in the relative stress graph were defined as true positive when a peak was also observed in the relative negative behaviour graph, and as false positive when this was not the case. Peaks in the relative negative behaviour graph that could not be matched with a peak in the relative stress graph were defined as false negative. This peak detection and matching resulted in an overall sensitivity of 77% and an overall precision of 78% for the entire dataset. This performance is regarded as sufficient to validate the methodology for stress detection in horses.

Two features were derived from the relative stress graph of every horse for each separate measurement. The first feature is the number of times the horse's relative stress exceeded a certain threshold during the measurement, further referred to as the #RT feature. The second feature is the time percentage during which a certain threshold was exceeded during the measurement, from here on referred to as the %RT feature. Both features were determined for five relative stress thresholds RT1, RT2, RT3, RT4 and RT5, respectively corresponding to 10, 20, 30, 40 and 50% of relative stress (Figure 4). The zones in which exceeding of the threshold occurs, are the stress-zones.

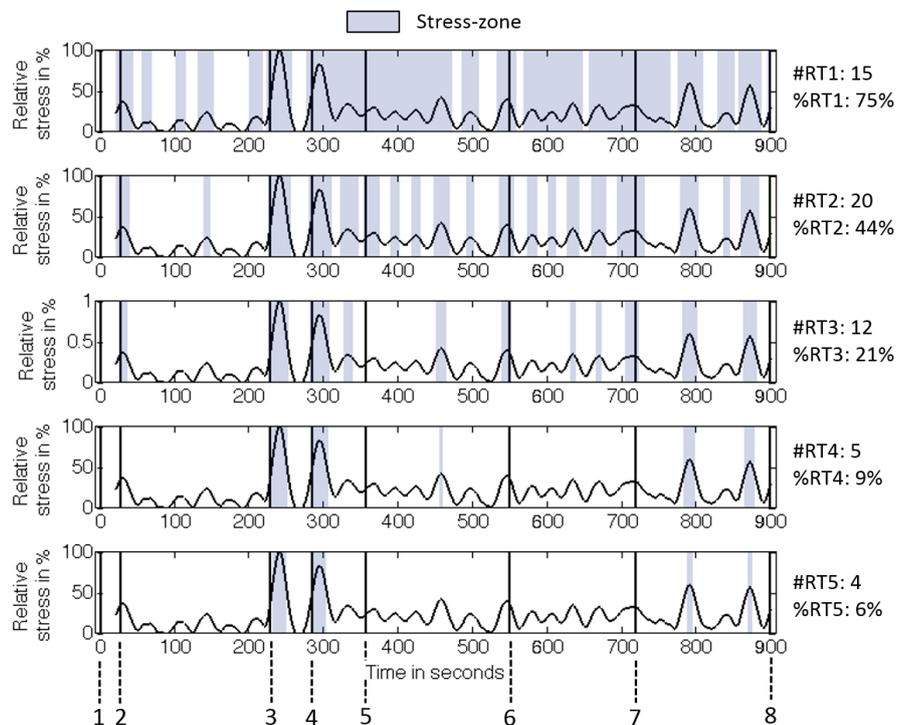


Figure 4: Feature extraction from the relative stress graph. The different steps in the protocol indicated in the figure are: 1) Standing still, 2) Walking, 3) Rider dismounts, 4) Attach horse, 5) Leave horse alone, 6) Prepare horse for box 7) Leave horse alone, 8) End.

3.6 Statistical analysis

For the comparison of the features between the four horse categories, a multiple comparison test was conducted. The feature values of all protocols were included in this comparison. Given the unequal sample sizes of the different categories, the Tukey-Kramer procedure was applied. In order to apply this test, the data of the different categories must be normally distributed and have equal variance (Rogan, et al., 1977; Somerville, 1993). First, the #RT and % RT feature values were tested for normality within each horse category using the Lilliefors test of normality with a significance level $\alpha = 0.05$ (Scott, et al., 2001). Then it was tested whether the features have equal variance in the four horse categories using the Bartlett's test for equal variances with a significance level $\alpha = 0.05$ (Ma, et al., 2015). The features with normal distribution and equal variance were further analysed using the Tukey-Kramer multiple comparison test with a significance level $\alpha = 0.05$

4 Results

4.1 Evaluation of the horse by the rider and the instructor

Summing the scores for the negative and positive statements resulted in a negative score (NS) and positive score (PS) for every horse, varying in the range of 0-10 and 0-30 respectively. Both the rider and the instructor completed the evaluation for every horse so that two positive and two negative scores are available per repetition of each horse. Complete evaluation by rider and instructor was available for 11 out of the 17 rider-horse pairs. The results are shown in Figure 5. A first observation is that for experienced horses the scores given by the instructor and rider are fairly consistent. Moreover, the scores of the experienced horses are logically grouped per category (GE and BE), in that sense that both categories are relatively well separated from each other in the graph. This is in contrast to the scores of the beginner horses, which show clear overlap. A second observation is that the rider and instructor can give very

different scores to the same horse, which illustrates the subjectivity of the evaluation. In addition, the finding that the evaluation of beginner horses lacks consistency confirms that an additional objective tool to evaluate suitability of police horses might prove useful for the mounted police.

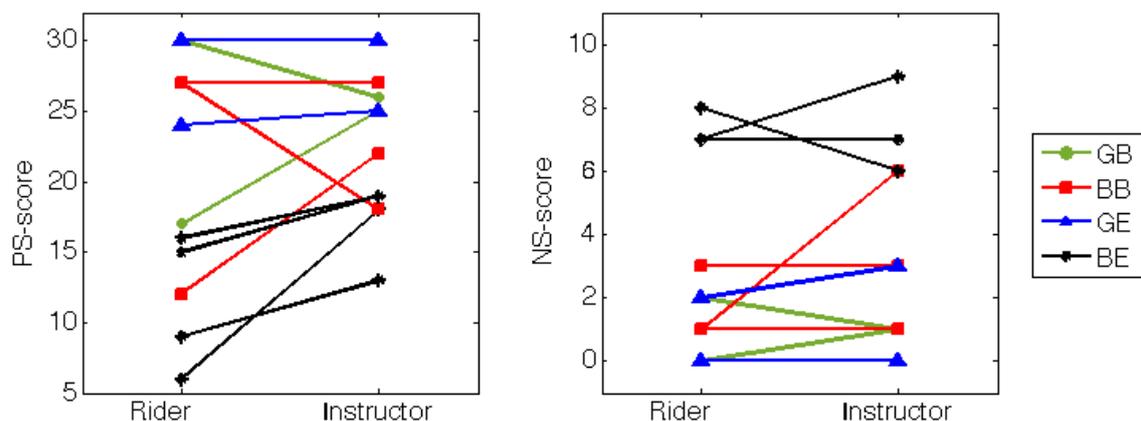


Figure 5: Positive Scores (PS) and Negative Scores (NS) given by the rider and the instructor to the horses from different categories. GB: Good Beginner – BB: Bad Beginner – GE: Good Experienced – BE: Bad Experienced

4.2 Stress of the horse

The features for which the four horse categories have normally distributed data according to the Lilliefors test are #RT1, %RT1 and %RT2 ($p > 0.05$). By means of the Barlett's test it was found that all three features have equal variance in the four horse categories ($p > 0.05$). Accordingly, the Tukey-Kramer multiple comparison test was carried out for these three features. It was concluded that none of the horse categories were significantly different from each other for any of the three features when combining the feature values of all protocols ($p > 0.05$). Figure 6 and Figure 7 visualise the 95% confidence intervals of the %RT1 and %RT2 values for the four horse categories. From these figures it becomes clear that there is a bigger difference in these feature values between GB horses and BB horses, namely that GB horses consistently show lower values for both features than BB horses. This might indicate that the time spent in certain stress-zones can be informative of the suitability of beginner horses.

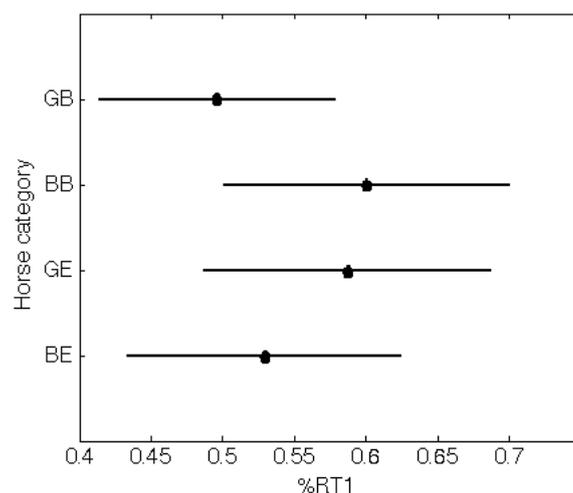


Figure 6: 95% confidence intervals for the %RT1 values (time percentage a horse exceeded the 10% relative stress threshold) of the four horse categories including all protocols. GB: Good Beginner – BB: Bad Beginner – GE: Good Experienced – BE: Bad Experienced

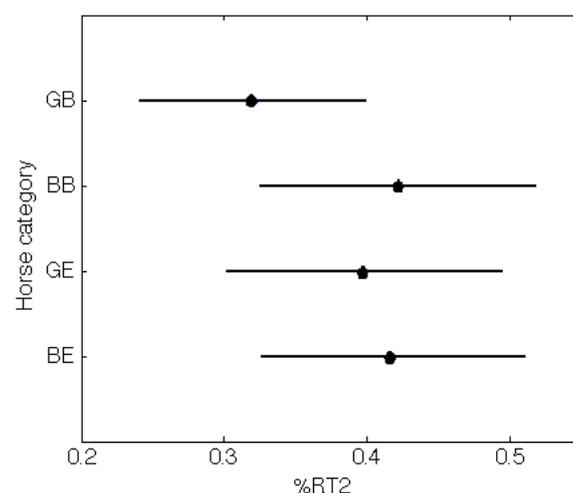


Figure 7: 95% confidence intervals for the %RT2 (time percentage a horse exceeded the 20% relative stress threshold) values of the four horse categories including all protocols. GB: Good Beginner – BB: Bad Beginner – GE: Good Experienced – BE: Bad Experienced

Comparison of the feature values between the different horse categories was also performed for each protocol separately, in order to investigate whether the feature values during a particular protocol might provide information on the suitability of a horse. The same statistical analysis was applied as previously described. Findings were that for the CFW, the values of the %RT2 feature are significantly lower for GB compared to BB ($p=0.0277$). In other words, during the CFW GB spend less time in the stress zones defined by RT2 compared to BB, which is in accordance with previous findings about the lower observed %RT2 values for GB horses compared to BB horses when evaluating all protocols in combination.

5 Discussion

The first objective of this study was to design experimental protocols that are representative of situations at the mounted police in Brussels. This objective was reached by developing four protocols together with the trainers and riders of the mounted police in Brussels. Not only did the protocols represent actual mounted police situations, they were also designed in such a way that they easily fit in the daily training schedule of the mounted police. Therefore the tests designed in this study can be conveniently implemented as a part of the training routine of the mounted police in Brussels in the selection and evaluation procedure for mounted police horses.

The second goal of this study was to find an objective measure that can help the mounted police to evaluate the suitability of mounted police horses. For this purpose the relative stress of the horses was calculated from heart rate and activity, measured by wearable technology. The method that is presented in this paper makes it possible to monitor the stress of police horses continuously and fully automatically in any situation. This is in contrast to previous studies where physiological and behavioural measures are either monitored in a discrete way or require a lot of manual work. For instance, in the study by Pierard et al. (2017), the measured variables, such as time spent trotting, have to be manually identified using video footage and specialised software. This has to be done every time the tests are run, which is labour intensive and unpractical for the mounted police. By monitoring the heart rate and heart rate variability of the horses using wearable technology, Munsters et al. (2013) recorded the physiological response of the horses in a fully automated way. Additionally, this wearable technology makes it possible to record data continuously. However, a lot of the information in these continuous data is lost when only the mean heart rate and mean heart rate variability are taken into account in the analysis. Moreover, the methodology presented by Munsters et al. (2013) only allows the comparison of heart rate and heart rate variability of activities of equal exercise intensity. The

current paper proposes a methodology where the intensity level of the activity is taken directly into account by measuring the acceleration of the horse at any time. Using this acceleration data, a calculated heart rate is obtained that only reflects changes in heart rate due to physical activity. Any additional measured heart rate can then be assigned to stress. This methodology was validated against the recorded behaviour of the police horses and was found to perform with a sensitivity of 77% and a precision of 78%. Once this methodology was validated, it could be used fully automatically for continuous stress measurement of police horses.

When observing all protocols, it was found that the time percentage spent above 10% and 20% relative stress can provide information on the suitability of beginner mounted police horses, although differences between the four horse categories were not significant. Observing only the CFW protocol, the %RT2 feature values were found to be significantly lower in GB horses compared to BB horses. The fact that GB horses spend less time in the 20% stress zone can either indicate that they experience less stress in general, but it could also mean that they recover more quickly from stress. The latter might explain why a significant difference could only be found in the CFW protocol, because in this protocol the horses are ending their training and are starting their physical and mental recovery. The CFW protocol is executed after the OT and the FT, which elicit the biggest stress responses in the horses. During the CFW protocol, no stress impulses are given to the horses and the horses are mainly standing still in the cleaning area. Therefore it is possible that stress that is detected during the last protocol is left-over stress from the previous protocols. Accordingly, horses that relax more quickly after a stress impulse would show less stress in the CFW protocol. It would be interesting to investigate this hypothesis further in a separate study.

Previous studies by Munsters et al. (2013) and Pierard et al. (2017) found that no correlation could be found between the experience of police horses and their behavioural or physiological response. That finding was confirmed in this study, as no significant differences

in relative stress could be found between beginner and experienced horses. According to Munsters et al. (2013) this could mean that experience as a police horse is not a key factor in how a police horse deals with stress.

This study has also shown that mounted police riders find it more difficult to evaluate new horses compared to known horses. This was illustrated by the lack of consistency in the evaluation the horses were given by different police riders, demonstrating the need for an objective tool to help in the selection and training of new police horses. To the authors' knowledge it is the first time that evaluation of police horses by the police riders themselves has been studied in such detail. In only one other study, conducted by Munsters et al. (2012), police riders were asked to assess the performance of their horses, but in that study the scores of the police riders were not analysed separately. However, as shown in this study, doing that type of analysis can be valuable. For example, we showed that despite years (3 to 30 years, with an average of 12 years) of horse riding experience, it is difficult for the police riders to assess the suitability of new police horses and that objective tools may be useful in selecting and training new police horses.

The evaluation of horses based on their relative stress can be valuable in other domains as well. For instance, when private individuals are looking for a new horse, they could use this tool to assess how a horse reacts to certain stimuli in order to make the best possible decisions when buying a horse. In professional settings a wearable stress monitor for horses could help riders and grooms to not only select certain types of horses. It could also be used to monitor them on the long term and keep an eye on the mental health of their horses, which is known to be a key factor for good performance. Finally this technology could also be applied in the field of animal welfare, for instance to evaluate if horses are housed, cared for and transported in the proper conditions.

Given that a relatively small sample size was considered here, it is advisable that further studies are carried out with a bigger number of horses. It would also be recommended to focus on beginner police horses in a follow-up study. A first reason is that this pilot study has shown that significant differences in relative stress can be found between bad and good beginner horses, which is not true for experienced horses. A second and more important reason is that facilitating the training and selection procedure of new police horses is of bigger interest for the mounted police.

6 Conclusion

This study is the first to have designed relevant and easily implementable experimental protocols to test the suitability of beginner police horses. In addition this study has shown promising results towards finding the first reliable objective measure for suitability assessment of police horses by measuring their relative stress using wearable technology. Although this study is limited in its sample size, it has demonstrated the potential of stress monitoring for suitability assessment at the mounted police, and potentially also in other domains.

7 Declaration of interest

The authors of this paper certify that there are no conflicts of interest.

8 Acknowledgements

The authors of this paper would like to thank the mounted police in Brussels for their collaboration and support throughout this study.

References

- Aerts, J. M., Gebruers, F., Van Camp, E., & Berckmans, D. (2008). Controlling horse heart rate as a basis for training improvement. *Computers and electronics in agriculture*, *64*(1), 78-84.
- Delanghe, B. (2014). Personal communication.
- Dictionaries, O. (n.d.). Retrieved from Oxford Dictionaries:
<http://www.oxforddictionaries.com/definition/english/equitation>
- Exadaktylos, V., Jansen, F., Van der Krogt, J., Guarino, M., & Berckmans, D. (2013). Real time monitoring of mental status of a horse to a negative stressor. *Precision Livestock Farming '13 ECPLF*. Leuven.
- Federale Politie. (n.d.). Retrieved 2014, from De politie te paard: http://www.polfed-fedpol.be/org/org_dga_cav_nl.php
- Flentje, R., & Creighton, E. (2010). Can standardized behavior tests predict suitability for use in horses. *Journal of Veterinary Behavior*, *5*(1).
- Forster, P., Hurles, M., Jansen, T., Levine, M., & Renfrew, C. (2012). Origins of the domestic horse. *National Academy of Science of the United States of America*, (pp. E3148-E3148).
- Hall, M., Frank, E., Holmes, G., Pfahringener, B., Reutemann, P., & Witten, I. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, *11*(1).
- Hausberger, M., Roche, H., Henry, S., & Visser, E. (2008). A review of the human-horse relationship. *Applied Animal Behaviour Science*, *109*(1), 1-24.

- Jansen, F., Van der Krogt, J., Van Loon, K., Avezzu, V., Guarino, M., Quanten, S., & Berckmans, D. (2009). Online detection of an emotional response of a horse during physical activity. *Veterinary Journal*, *181*(1), 38-42.
- Keeling, L., Jonare, L., & Lanneborn, L. (2009). Investigating horse-human interactions: the effect of a nervous human. *Veterinary Journal*, *181*(1), 70-71.
- Levine, M. (1999). Investigating the origins of horse domestication. *Equine Veterinary Journal*, *31*(S28), 6-14.
- Ma, X., Lin, F., & Zhao, Y. (2015). An adjustment to the Bartlett's test for small sample size. *Communications in Statistics-Simulation and Computation*, *44*(1).
- Matlab. (n.d.). R2014a(8.3.0.532).
- Mengoli, M., Pageat, P., Lafont-Lecuelle, C., Monneret, P., Giacalone, A., Sighieri, C., & Cozzi, A. (2014). Influence of emotional balance during a learning and recall test in horses (*Equus caballus*). *Behavioural Processes*, *106*, 141-150.
- Munsters, C., Visser, E., van den Broek, J., & van Oldruitenborgh-Oosterbaan, M. (2013). Physiological and behavioral responses of horses during police training. *Animal*, *7*(5).
- Munsters, C., Visser, K., van den Broek, J., & Sloet van Oldruitenborgh-Oosterbaan, M. (2012). The influence of challenging objects and horse-rider matching on heart rate, heart rate variability and behavioural score in riding horses. *The Veterinary Journal*, *192*, 75-80.
- Peeters, M., Closson, C., Beckers, J., & Vandenheede, M. (2013). Rider and Horse Salivary Cortisol Levels During Competition and Impact on Performance. *Journal of Equine Veterinary Sciences*, *33*(3), 155-160.

- Pierard, M., & Geers, R. (2014). Responses of horses to sudden object and sudden noise in relation to direction and distance of the stimulus. *10th International Equitation Science Conference*, (p. 29). Bredsten, Denmark.
- Pierard, M., McGreevy, P., & Geers, R. (2017). Developing behavioral tests to support selection of police horses. *Journal of Veterinary Behavior*, 19, 7-13.
- Pierard, M., Verbeke, G., & Geers, R. (2013). The use of temperament tests for the evaluation of police horses. *9th International Equitation Science Conference*, (p. 49). Delaware, USA.
- Piette, D. (2015). *Real-time monitoring van mentale toestanden van paarden en ruiters*. KU Leuven: Master's Thesis.
- Police Nationale. (2017). Policier cavalier en unité équestre, Retrieved from: <http://www.lapolice.nationale.recrute.fr/Fiches-metiers/Policier-en-brigade-equestre>
- Rawdon-Mogg, J. (2016). An Introduction to the Household Cavalry Mounted Regiment. Retrieved online from http://www.army.mod.uk/documents/general/ADR006136_House_Cav_web.pdf#search=cavalry
- Rogan, J., Keselman, H., & Breen, L. (1977). Assumption violations and rates of type-I error for Tukey multiple comparison test – Review and empirical-investigation via a coefficient of variance variation. *Journal of Experimental Education*, 46(1).
- Royal Newfoundland Constabulary (n.d.). Mounted Unit. Retrieved (2017) from: <https://www.rnc.gov.nl.ca/about-us/what-we-do/mounted-unit/>
- Sankey, C., Richard-Yris, M.-A., Leroy, H., Henry, S., & Hausberger, M. (2010, February). Positive interactions lead to lasting positive memories in horses, *Equus caballus*. *Animal Behaviour*, 79.

Scott, W., & Stewart, B. (2001). Tables for the Lilliefors and Modified Cramer-von Mises Tests of Normality. *Communications in Statistics-Theory and Methods*, 40(4).

Somerville, P. (1993). On the conservatism of the Tukey-Kramer multiple comparison procedure. *Statistics & Probability*, 16(5).

Taylor, C., Pedregal, D., Young, P., & Tych, W. (2007). Environmental time series analysis and forecasting with the Captain toolbox. *Environmental Modelling & Software*, 22(6).

Thomas, R. (2010, July/August). Predictability in an unpredictable environment: Training the police horse using learning theory. *Journal of Veterinary Behavior*, 5(4).

Valenchon, M., Lévy, F., Fortin, M., Leterrier, C., & Lansade, L. (2013, November). Stress and temperament affect working memory performance for disappearing food in horses, *Equus caballus*. *Animal Behaviour*, 86.

Viazi, S., Ismayilova, G., Sonoda, L., Oczak, M., Leroy, T., Costa, A., . . . Berckmans, D. (2011). Labelling of video images: the first step to develop an automatic monitoring tool for pig aggression. *15th ISAH Congress*, (pp. 751-754). Vienna.

Visser, E., Van Reenen, C., Zetterqvist Blokhuis, M., Morgan, E., Hassmén, O., Rundgren, T., & Blokhuis, H. (2008). Does Horse Temperament Influence Horse–Rider Cooperation? *Journal of Applied Animal Welfare Science*, 11(3), 267-284.

von Borstel, U., Duncan, I., Shoveller, A., Millman, S., & Keeling, L. (2007). Transfer of nervousness from competition rider to the horse. In *Conference Proceedings 3d International Equitation Science Conference* (p. 16). East Lansing, MI USA: Michigan State University.

von Borstel, U., Euent, S., Graf, P., Konig, S., & Gauly, M. (2011). Equine behaviour and heart rate in temperament tests with or without rider or handler. *Physiology & Behavior, 104*(3).

von Lewinski, M., Biau, S., & Erber, R. (2013). Cortisol release, heart rate and heart rate variability in the horse and its rider: Different responses to training and performance. *Veterinary Journal, 197*(2), 229-232.

Watson, M., Lyon, R., & Montgomery, S. (2001). *Compleet handboek paarden*. Aartselaar, België: Deltas.

Wolframm, I., Bosga, J., & Meuleunbroek, R. (2013). Coordination dynamics in horse-rider dyads. *Human Movement Science, 32*, 157-170.