Impact of Training Method on Behavioral, Physiological, and Relationship Measures in Horses

Lindsay T. Isernia

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Animal and Poultry Sciences

Erica N. Feuerbacher (Chair)
Robin L. Foster
Timothy J. Jarome

(December 16, 2020)
Blacksburg, Virginia

Keywords: (Riding horses, rehomed Thoroughbreds, positive reinforcement, negative reinforcement, combination reinforcement)

Impact of Negative and Combination of Both Positive and Negative Reinforcement Training on Behavioral, Physiological, and Relationship Measures in Horses

Lindsay T. Isernia

ACADEMIC ABSTRACT

With a rise in concern for animal welfare, the equine world has started using positive reinforcement (R+); as such, horses often experience a combination of negative reinforcement (R-) and R+. I compared the effects of R- to a combination of positive and negative reinforcement (R-/R+) training. Horses were trained to walk across two visually discriminable liverpools (striped, Experiment 1; colored water, Experiment 2), each associated with either R- or R-/R+, and training type alternating across six days. I measured highest training criteria reached, prevalence of undesirable behaviors, salivary cortisol (pre- and post-training), time spent by the trainer in motionless human tests (pre- and post-training), and horses' preference for the two liverpools using concurrent choice. Across both experiments, I found no significant difference in the proportions of criteria reached between training types; horses engaged in mugging for longer periods of time in R-/R+ than R-; no significant difference between training types for the pre- to post-change of cortisol; a greater proportion of horses increased time spent with R-/R+ trainer than the R- trainer; and no difference between first choice in the preference test or time horses spent in proximity to the liverpool, based on the training type with which the liverpool was associated. Overall, I found few differences between R-/R+ and R-, which could be due to horses only having 30 min total training contact with either training, or my use of relatively low intensities of R- and R+.

Impact of Negative and Combination of Both Positive and Negative Reinforcement Training on Behavioral, Physiological, and Relationship Measures in Horses

Lindsay T. Isernia

GENERAL AUDIENCE ABSTRACT

The equine world has started using positive reinforcement (R+), such as providing treats. Often horses experience a combination of negative reinforcement (R-) and R+, such as having rein pressure released and being given a treat. I compared effects of R- to a combination of positive and negative reinforcement (R-/R+) training. Horses were trained to walk across two visually distinct liverpools, a 1 m X 2.7 m shallow pool, (striped, Experiment 1; colored water, Experiment 2) each associated with either R- or R-/R+, and training type alternating across six days. I measured highest training level reached, occurrence of undesirable behaviors, salivary cortisol (a measure of stress), time spent by the trainer in motionless human tests, and horses' preference for the two liverpools. Across both experiments, I found no significant difference in the proportions of criteria reached between training types; horses investigated the trainer for treats for longer durations in R-/R+ than R-; no significant difference between training types for the pre- to post-change; a greater proportion of horses increased time spent with R-/R+ trainer than R- trainer; and no difference between first choice in the preference test or time horses spent in proximity to the liverpool, based on the training type with which the liverpool was associated. Overall, I found few differences between R-/R+ and R-, which could be due to horses only having 30 min total training contact with either training, or my use of relatively low intensities of R- and R+.

DEDICATION

I would like to dedicate this thesis to all my family and friends that believed in me and encouraged me to reach my goals. I love you all.

ACKNOWLEDGMENTS

My Family. Thank you for always being a wonderful support system, for being there when times were hard, and pushing me to reach my goals.

Dr. Feuerbacher, Dr. Foster, and Dr. Jarome. Thank you all for being on my committee and for your support as I completed my thesis.

JoAnna Platzer and Hannah Decker. Thank you for your countless hours of help even though you are both extremely busy graduate students!

Research Assistants. Thank you all for the many hours you put in to help me collect data and code videos.

This work was supported in part by the Paul Mellon Research Endowment.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

History of the Horse Industry

Interactions between horses and humans started 15,000 years ago (Waran et al., 2007). Shortly after the start of those interactions, humans began to handle horses and about 9,000 years later began to ride them (Waran et al., 2007). During that time period, there is strong evidence that shows that horses were handled and trained using negative reinforcement (R-) (Waran et al., 2007). Today, the horse industry largely relies on R-. However, positive reinforcement (R+) techniques are becoming more common.

Negative Reinforcement

R- is defined as an increase in the frequency of a behavior that results from a withdrawal or termination of a stimulus(Cooper et al., 2014). Most training aids in horse training and riding rely on the principle of R-. For example, if a handler applies pressure on the horse's halter and contingent removal of the pressure when the horse steps forward results in the horse being more likely to step forward in the future when halter pressure is applied, the removal of the pressure would be R-. Another example would be if a rider applies pressure with the side of their leg and contingent removal of the pressure when the horse jogs results in the horse being more likely to jog in the future when leg pressure is applied, the removal of the pressure would be R-

While poor timing of reinforcement can inadvertently reinforce the wrong and often undesirable behavior, poor timing in R- goes beyond the problems of poor timing with R+. First, because R- relies on applying an aversive stimulus that can then be contingently removed, if handlers do not remove the aversive stimulus at the appropriate time that is contingent on the desirable behavior, the application of the aversive stimulus for a long duration can increase the likelihood of aversive-induced aggression. Even worse, if the handler now releases pressure when the undesirable, possibly aggressive, behavior occurs, that behavior will be negatively reinforced and increase in frequency. For example, if a rider applies leg pressure for the horse to move forward and the horse moves forward, but the rider does not remove the pressure, the horse's behavior of moving forward is not reinforced by contingent removal of the leg pressure and could even be punished by continued application of the aversive stimulus (McGreevy & McLean, 2009). Instead, if the continued application of the leg pressure results in the horse engaging in another behavior that causes the rider to lose their balance and fall off (e.g., bucking), the rider falling off could inadvertently reinforce the bucking behavior because thel pressure was released when the rider fell (Jones, 2017). Furthermore, if the pressure is not released at the appropriate time it can cause the horse to habituate to pressure and the horse can become dangerous (McGreevy & McLean, 2007).

Positive reinforcement

Recently, more R+ is being used and explored as a training tactic in the equine industry. According to Cooper et al. (2014), positive reinforcement (R+) is an increase in frequency of a behavior that resulted from an added stimulus in the environment. For example, if a handler is training a horse to load in a trailer, the handler can deliver a handful of grain each time the horse takes a step towards the trailer. If the contingent delivery of grain increases the likelihood that the horse steps toward or into the trailer in the future, the delivery of grain would be positive reinforcement.

A variety of stimuli have been evaluated as possible positive reinforcers for horses, including grooming (Sankey, Henry, et al., 2010), scratching (Kieson et al., 2020) and patting (Olczak et al., 2020). Nevertheless, food has typically been the most effective positive reinforcer for horses (Kieson et al., 2020; Olczak et al., 2018; Williams et al., 2004). Kieson et al. (2020) compared food and interactions with humans as reinforcement. They trained horses to touch three different symbols with their muzzle. If the horse touched the "X" with its muzzle it was given food. If the horse touched the "O" it was given scratches by the human. If the horse touched the solid square, it was given patting by the human. Once horses had associated each symbol with the correct consequence, they let the horse choose a symbol and measured their preference. All horses preferred food reinforcement over human interaction, indicating that food might be a more effective reinforcer than scratching or patting.

This conclusion was corroborated by Sankey, Henry, et al. (2010) in which they evaluated how horses performed with food versus grooming as a consequence. They found that horses that were given food reinforcement were able to complete the training task (standing still) more quickly and stayed immobile for longer periods of time than horses that were given grooming as reinforcement. This demonstrates that, compared to other interactions, food is a more effective form of reinforcement for horses.

Nevertheless, the amount of responding that a given food reinforcer will maintain does vary across horses. Olczak et al. (2018) investigated horses' food motivation through two tests: pulling a lever to receive food with an increasing work ratio and crossing a barrier to get to the food reward, in which the barrier height increased with each trial. Horses varied in the amount of behavior that they engaged in: Some horses worked for a longer period of time while others worked for shorter period of time. For example, in the lever test, one horse pressed the lever 21 times before receiving reinforcement, where three other horses only pressed the lever only 3 times. For crossing a barrier, one horse reached a level of 90 cm whereas four horses only reached a level of 55 cm. This suggest that food motivation may vary across individuals. Horses, like dogs, likely vary in how sensitive they are to food reinforcement and how much responding it will maintain (Feuerbacher & Wynne, 2012; Vicars et al., 2014). The exact food type that is most reinforcing can also differ across horses and Olczak et al. (2018) did not conduct a preference test prior to determine the individual horses' preferred food. This might account for some of the individual variation they saw in this study. However, they did find a correlation between horses that pressed the lever multiple times and horses that crossed the barrier at multiple heights, suggesting that the individual variation in sensitivity to food reinforcement transcended the specific task the horses were working on. Additionally, horses that engaged in more behavior, also ate the food reinforcement faster, suggesting a general increased motivation for food in those horses.

Behavioral Side Effects of R- and R+ Training

While the main effect of R+ and R- is to increase in frequency the behavior that preceded the reinforcer, and both are useful for training horses, these two reinforcement strategies differ in their behavioral side-effects. These side-effects should impact which type of training we choose to use.

R- requires the application of an aversive stimulus, which has been shown to induce aggressive behaviors. Azrin et al. (1963; 1964) conducted a series of studies looking at fighting reactions in several species when a shock was delivered. For example, when rats were shocked, they almost always engaged in fighting. The frequency of these aggressive behaviors also increased in frequency when an aversive stimulus was repeatedly applied. Azrin et al. (1963) placed two monkeys in an experimental space and delivered brief shocks to them every 30 sec for 5 min. As the number of shocks increased, the amount of fighting increased; the first shock seemed to startle the first pair of monkeys, but the third shock resulted in the monkeys fighting long after the termination of the shock. This study suggests that application of an aversive stimulus and the frequency of application can lead to increased aggression. A survey conducted by Blackwell et al. (2008) found that aggression was seen significantly more frequently when owners used positive punishment. This needs to be kept in mind when considering using an aversive stimulus to train horses, which can weigh upwards of half a ton, and could be one reason why they sometimes show aggressive behaviors.

The intensity of the aversive stimulus or environmental factors can also induce aggressive behaviors. Azrin et al. (1964) investigated whether changing circumstances impacted reflexive fighting. They investigated the impact of the intensity of shock on the subject's behavior towards an inanimate object and other animals. Monkeys were placed into the experimental space with either a ball, a doll, a mouse, a rat, or another monkey. Eleven out of 14 monkeys attacked another monkey or rat at least once, and the monkeys also fatally injured the mice. The researchers determined that the probability of attack increased with the intensity of the shock. This study suggests that the application of aversive stimuli in animal training, as needed to use R-, can potentially induce aggression, and can be particularly likely with high intensity aversive stimuli. For example, when lunging a horse, the use of pressure from a whip might be applied to get the horse to move forward and contingently removed to negatively reinforce the forward movement. However, if the person waves the whip harder or faster, this could increase the likelihood of the horse to engage in defensive behaviors, such as striking out, biting or kicking (Sankey, Richard-Yris, Henry, et al., 2010). Those behaviors can lead to an unsafe environment for both the trainer and the horse, and further, if those behaviors result in the horse escaping the aversive stimulation, those behaviors will be negatively reinforced and increase in frequency in the future. When using R- with horses, using the lowest level of aversive stimulation possible will reduce the likelihood of aggression.

While the use of aversive stimuli can induce aggression, the use of appetitive stimuli in R+ can potentially decrease problem behaviors. Sankey, Richard-Yris, Henry, et al. (2010) found that during training, horses trained with R+ displayed more positive behaviors such as sniffing and licking and fewer negative behaviors, such as biting or kicking than horses that did not receive positive reinforcement. When working with horses that were problem loaders and previously trained with R-, Ferguson and Rosales-Ruiz (2001) reported that problem behaviors decreased to zero when they began training with R+ using target training and shaping. Additionally, while R- produces escape and avoidance behaviors, R+ can produce approach behavior, which are often measured in approach to humans in motionless tests. These will be discussed in a subsequent section.

Because R- requires the application of an aversive stimulus and R+ is the delivery of an appetitive stimulus, these can also produce differential side-effects through classical conditioning. Classical conditioning occurs when an animal makes an association between two stimuli, one which was previously neutral (NS) and another stimulus that currently elicits a behavioral response from the animal, often an unconditioned stimulus (US) that elicits a reflexive behavior (McLean & Christensen, 2017). Through contiguous and predictive pairings between the NS and US (Rescorla, 1967) the NS begins to elicit the same or similar response that the US did; in this way the NS becomes a conditioned stimulus (CS). While a trainer might use R- or R+ for the operant effects of increasing the frequency of a behavior, classical conditioning is also occurring. If a handler uses R+, the handler which was an NS for the horse might now become a CS through being paired with food. This means that the horse would likely show approach behaviors to the trainer; this has been borne out in post-training human-horse interaction tests (Sankey, Henry, et al., 2010; Sankey, Richard-Yris, Henry, et al., 2010; Sankey, Richard-Yris, Leroy, et al., 2010). On the other hand, if the trainer uses R-, the trainer could become a CS through pairings with the aversive stimulus. The trainer could then evoke the same behaviors as an aversive stimulus, such as escape, avoidance, and even aggression. In a study which compared dogs receiving shock collar training to other training, those that had received shock from their owners showed more stress signals than dogs that did not receive shock and this occurred even when the owner was present or giving commands in the absence of shock (Schilder & van der Borg, 2004). That is, the owner and their cues had become conditioned aversive stimuli for the dogs. The same conditioning can be seen in horses; horses, in fact, have shown behavioral and neurophysiological differences to human voices associated with positive (food) or negative (inedible, vinegar-soaked food) experiences (d'Ingeo et al., 2019).

Measures of Welfare

Possible differences in R- and R+ can also be assessed by measuring their impacts on markers of stress, which can be indicative of the animal's welfare. Welfare indicators can include behavior such as ethological observations and cognitive bias tests, as well as more physiological measures such as cortisol levels and heart rate.

Behavioral indicators

Not only are the differential behavioral side-effects outlined above between R+ and R-, with possibly more escape, avoidance, and aggression occurring with use of R-, these behaviors might also indicate increased stress levels and reduced welfare for the horse. Watson and McDonnell (2018) suggested that behaviors such as pawing, stamping, tail and head movements, vocalizations, and defecation can all indicate stress in horses and that these can be produced from mild aversive stimulation (in this study, a mildly noxious auditory stimulus), which could include traditional aversive stimuli used in R- training.

R+ on the other hand delivers appetitive stimuli, such as food, contingent on the desired behavior. Fewer problem and stress-related behaviors are observed when R+ is utilized. For example, Dai et al. (2019) investigated the effect target training followed by a shaping procedure using R+ had on a group of meat horses during trailer loading, which can be one of the most stressful situations when it comes to animal transport (Padalino et al., 2012). Horses were trained to touch a target stick with their muzzle and then their behavior was shaped until they followed the target into the trailer. Dai et al. (2019) found that horses who were trained with target training took significantly less time and showed more forward movement to load than the control group, which received no training. While they did not find significant differences in durations of stress-related behaviors, the authors did note that the stress-related behaviors they did observe (e.g., kicking, rearing, pawing, defecating) occurred only in the control group. Although this study did not include a sham-handled group of horses to ensure this behavioral difference was due to the specific training and not simply increased time being handled by the human, the results from Ferguson and Rosales-Ruiz (2001) suggest that the switch to R+ can produce reductions in undesirable and stress-related behaviors. They found that when R+ was started during trailer loading training, undesirable behaviors (e.g., freezing, rearing, head tossing) fell to zero. This further supports the use of R+ as an effective training technique to reduce stress-related behaviors, and might be especially useful in stress-inducing situations.

Another way of assessing welfare through behavior is by using cognitive bias tests. Clegg (2019) defines cognitive bias testing as a way to "measure how emotional states can affect cognitive processes” (p. 1). Freymond et al. (2014) used a bias test to evaluate horses’ emotional state between using R+ versus R-. They found that horses trained with R+ traveled significantly more slowly to the negative location than horses trained with R-. They concluded that R- horses may have traveled faster to the negative location because they were used to being trained with R- and therefore the contrast with R+ and food delivery was greater or because the R+ horses were more satiated on food. They also looked at the horses’ behavior during training sessions. When interacting with the trainer, R+ horses showed more motivation and focus, with ears forward, and a lower headset than the R- mares (Freymond et al., 2014). These are all taken as evidence of the horse having more positive emotions; ears forward can be a visual cue of attention (Wathan & McComb, 2014) and horses carrying their head below their withers can be a sign of relaxation (Thorbergson et al., 2016).The behaviors seen in the Freymond et al. (2014) study during R+ training suggest that R+ can be a product of training and horses are more motivated to work when receiving R+.

Cortisol

Cortisol is a common biological marker for assessing stress in animals (Beerda et al., 1996; Desportes et al., 2007; Glenk et al., 2013). Cortisol levels can be measured through saliva, urine, or plasma. However, when assessing stress levels in animals, Beerda et al. (1996) and Cook (2012) found that measurements like blood sampling are relatively invasive and alone may affect stress levels. Thus, salivary cortisol sampling might be a less invasive, and preferred, approach (Beerda et al., 1998). Additionally, salivary cortisol is a sensitive indicator of adrenocortical activity compared to total concentrations found in plasma (Valera et al., 2012).

Salivary cortisol samples have been found to reflect the application of an aversive stimulus. Beerda et al. (1998) found that when dogs were presented with an aversive stimulus, salivary cortisol sampling was able to help capture their stress level. In horses, Arnell et al. (2013) found that salivary cortisol concentrations increased after clipping, which is a potentially aversive event, compared to before the clipping procedure. This study suggest that saliva cortisol can be used to determine if a procedure is aversive to a horse.

Salivary cortisol sample have also been used in assessing R+ or appetitive training techniques. As discussed before, Olczak et al. (2018) looked at horses' food motivation when presented with a lever test and crossing a hurdle. Along with looking at behavioral effects, they collected saliva cortisol samples pre- and post-testing. Samples showed little difference in pre- and post-testing results, suggesting that the training and testing might not have been stressful. This study shows that salivary cortisol sampling can also be successful when assessing R+ training methods and suggests that R+ training might not have the same effect on cortisol levels as the application of an aversive.

Salivary cortisol sampling can also be helpful when assessing stressors unrelated to training when using R- or R+ in horses (Valenchon et al., 2017). Valenchon et al. (2017) divided horses into four groups, R- without an unrelated stressor, R- with an unrelated stressor, R+ without an unrelated stressor, and R+ with an unrelated stressor. Saliva samples were collected pre- and post- learning procedures in each condition for Trials 1 and 3. When an unrelated stressor was not used, there was no statistical difference in cortisol levels between R- and R+. However, when an unrelated stressor was added, there was a statistical increase in cortisol for R+ compared to R-. The results of the salivary cortisol mirrored the impact of stressors unrelated to training on horses' ability to learn, with the unrelated stressor impacting learning in both R+ and R-, but with a much larger impact in the horses being trained with R+. This study further points out the utility of salivary cortisol for assessing stress in horses and that it can parallel behavioral observations.

Heart Rate

Heart rate can also be an indicator of emotional reactions in both R+ and R- training (Herd, 1991; Jezierski et al., 1999; Visser et al., 2002) and might be affected differently by R+ and R- training. Sankey, Richard-Yris, Henry, et al. (2010) compared heart rate in horses that were trained to back up with R+ or R-. The R+ group was given a vocal command followed by food reinforcement and the R- group was given the vocal command followed by using and then removing a whip. They discovered that horses who were trained with R- had a higher heart rate from the time the first vocal command was given to the end of the training session compared to horses trained with R+ (Feh & de Mazières, 1993). Similarly, Feh and de Mazières (1993) found that when they imitated grooming that happens between two horses in the wild (an appetitive stimulus), horses' heart rates decreased. On the other hand, Yarnell et al. (2013) also reported that when horses interacted with aversive stimuli, such as clipping, heart rate significantly increased during the first five minutes of clipping. Both studies suggest that when presented with an aversive stimulus, like a whip or clippers, there in an increase in the horse's heart rate.

While these studies suggest that changes in heart rate might be able to be used to determine whether a horse is expeirencing an appetitive or aversive stimulus, this is not always the case and it is most effectively used in combination with another measurement, like behavioral observations. McDonnell and Watson (2018) found that while wither scratching, face and eye rubbing, and food reinforcement all produced decreased avoidance and stress behaviors, they did not have a significant effect on heart rate, although McBride et al. (2014) reported that scratching the withers did decrease heart rate. Changes in heart rate appears to be a potentially useful measure for assessing the valence of different stimuli for horses, although more studies are needed to substantiate whether changes in heart rate are predictive of aversive or appetitive stimulation. Nevertheless, the results across a variety of behavioral and physiological stress measures suggest that R+ might produce better welfare in horses than R-.

Relationship impacts

Finally, R+ and R- might produce differential outcomes on the human-horse relationship. As discussed earlier, these different training techniques can differentially condition the trainer to be a conditioned appetitive or a conditioned aversive stimulus. This conditioning is often assessed through the motionless human test and is thought to measure aspects of the horse-human relationship. This test involves an experimenter standing in the middle of the testing arena to see how long the horse interacts with the experimenter, the behaviors the horse exhibits toward them, and the latency until the horse begins to interact with the experimenter (Sankey, Henry, et al., 2010; Sankey, Richard-Yris, Henry, et al., 2010; Sankey, Richard-Yris, Leroy, et al., 2010). In all three studies the results showed that horses that received R+ training spent a greater amount of time with the experimenter and had a shorter latency to approach the experimenter than horses that were trained with R- or no reinforcement.

This has also been assessed with large intervals of time between the training and the motionless human test. Sankey, Richard-Yris, Leroy, et al., (2010) investigated whether horses trained with R+ or horses that did not receive R+ during training would show differential responding to their trainer when tested immediately or six months after training. They also evaluated whether any pro-social behaviors that the horses showed to their trainer would generalize to a stranger by testing the horses with a stranger eight months after training. At each time point, they conducted the standard motionless person test and an approach contact test (experimenter stands 2 m away from horse and walks towards the horse's shoulder at a 90 degree angle; latency to contact was measured).

Horses trained with R+ spent more time in contact with the trainer in the motionless human test immediately after training and also at the six-month testing timepoint than the horses that did not receive R+ (Sankey, Richard-Yris, Leroy, et al., 2010). Additionally, this pro-social behavior to the trainer generalized to a stranger: when tested eight months after training, horses trained with R+ spent more time in contact with the stranger in the motionless human test immediately after training than the horses that did not receive R+. Horses trained with R+ were also faster to accept human contact on the shoulder at all test points (from both the familiar trainer and the unfamiliar stranger) than the horses that did not receive R+. This study provides strong evidence that the use of R+ training can produce more pro-social behaviors in horses towards humans, and impact the human-horse relationship. Interestingly, not only did R+ training impact the horses' behavior towards their trainer, it seemed to generalize and produce pro-social behaviors in the horses towards strangers as well.

The impact of training techniques on animal-human interactions in horses is corroborated by results found in dogs. Deldalle and Gaunet (2014) reported that dogs trained with R+ gazed at their owner longer than dogs that were trained with R-. Elliot and King (1960) reported puppies that had been fed by a human in the past exhibited more affiliative and fewer avoidance behaviors than puppies that had not been fed by a human. These results suggest that R+ training produces more affiliative behaviors in dogs toward humans than does R- training. The combined results of studies looking at the effects of R+ on animal-human interactions argue for an increased use of R+ in horse training, as it can potentially improve horses' interactions with their owners.

Using a Combination of R- and R+

While plenty of studies have looked at just R +, just R-, or R+ versus R- (e.g., Innes & McBride, 2007; Dai et al., 2019; Fenner et al., 2019) there is less research that compares R- and a combination reinforcement (using R-/R+ concurrently). Because many horses have started their training in R- and might have had R+ added on later and experience both R- and R+ during training, this is an important area to investigate. McGreevy and Warren-Smith (2007) compared using R- to a combination of both R+ and R- to train a horse to halt. Horses that were in the R- group were trained to halt by applying tension on the long reins and releasing the tension once the horse halted. Horses that were in the R-/R+ group were trained to halt through the same release of rein tension plus the delivery of molasses water through their bit once the horse halted. Horses that were trained with the R-/R+ shook their head vertically less than the horses trained with only R-; unsteady head carriage, such as vertically shaking their head, can indicate discomfort (McGreevy & Warren-Smith, 2007). Horses in the R-/R+ training also licked their lips more frequently than horses trained with only R-; licking can be perceived as a positive affiliation (Sankey, Richard-Yris, Leroy, et al., 2010), such that this could indicate that R-/R+ training could improve horses' welfare.

Heleski, Bauson, and Bello (2014) also investigated the use of R-/R+ versus R-. Horses were divided into a R-/R+ group, which included applying and releasing pressure on the halter and delivery of oats, and a R- group, which only had applying and releasing pressure on the halter. All horses were trained to walk over a tarp in a ten-minute session and were scored on a 1-5 scale, with 1 denoting that the horse went over the tarp with little hesitation and 5 denoting that the horse was extremely resistant. Nine out of 34 horses failed to walk over the tarp within the ten-minute session. Of those nine horses, six were from the R- group and three were from the

R-/R+ group, suggesting that the addition of R+ could increase success when training a challenging task.

While these studies have evaluated some aspects of R-/R+, like behavioral measurements such as lip licking and head tossing and the time and level of difficulty it takes for a horse to complete a training task, they did not explore some of the other measures found to be useful when assessing the effects of training techniques on horses, specifically the effects of combination reinforcement on physiological measures, preference measures, and human-horse relationship measures.

In my study, I compared training horses to walk over a liverpool using R- to R-/R+. I evaluated the effects on behavioral, physiological, and relationship measures in horses and used a within-subjects design so that I could compare each horse's performance with R- to its performance with R-/R+, thus eliminating the possibility of individual differences effects. In both experiments, I trained horses to walk across the long axis of a visually discriminable liverpool. One liverpool was associated with R-/R+ and the other with just R- training, counterbalanced across horses. I evaluated training criteria reached as well as quantified the frequency or duration of observed behaviors. I collected saliva pre- and post-training for cortisol analysis, to determine whether training type impacted cortisol levels. Before and after each training session, I conducted a motionless human test to see whether training type impacted the human-horse relationship. Finally, after all training sessions had been completed, I tested horses' preference for the two stimuli used in training by allowing them concurrent access to both liverpools and measuring time spent with each. This test was to evaluate whether preference for one training type or the other could be determined. Preference tests have been used effectively in horses to allow them to indicate preferences for blanketing (Mejdell et al., 2016), and in dogs to indicate preference for different human interactions (Bonne Beerda et al., 1998)

MATERIALS AND METHODS

All procedures involving animals were approved by the Virginia Tech Institutional Animal Care and Use Committee (protocol #19-022).

Subjects

In Experiment 1, I tested 10 horses from the Virginia Tech Hunt Seat Lesson Herd (see Table 1a for horse demographics).

In Experiment 2, I tested 10 horses from the Virginia Tech Hunt Seat Lesson Herd (see Table 1b for horse demographics) and eight rehomed Thoroughbreds from the Virginia Tech Research Herd (See Table 1c for demographics). Groups were matched by sex but otherwise randomly assigned to experimental group. Two horses were dropped from the study: Artimus on the second day of training for stifle injections and Sondheim on the fourth day of training due to behavior concerns (rearing, pacing, neighing, and increased defecation while in the stall).

Experimental Schedule

The overall and daily experimental schedules are shown in Figure 1 and Table 2, respectively.

Overall Schedule

Each experiment occurred over the course of seven days: three days of R-/R+ training, three days of R- training, and one day for preference tests. For horses in Experiment 1 and some horses in Experiment 2 their first experimental day was R-/R+ training and their second day was R- training (see Table 1 for initial training condition of each horse); for the remaining horses in Experiment 2, their first experimental day was R- training and their second day was R-/R+

training. Type of training alternated each day for each horse until each horse completed three days in each condition.

Daily Schedule

Horse's began their day with a pre-training motionless human test, followed by two minutes of rest. I then collected a pre-training salivary cortisol sample, after which horses completed their 10 min training session and were placed back into their stall for a 30 min wash-out period. After the washout period, I collected the horse's post-training salivary cortisol sample and gave them a two-minute break. The horses then performed their post-training motionless human test and were placed back into their stalls and given hay.

Training

For Experiment 1 and 10 horses in Experiment 2, horses were tested at Virginia Tech's Campbell Arena, a covered arena with sand footing. On one end of the arena, I constructed a round pen (16.76 m diameter) from fourteen 3.66-m panels (Figure 2a) for use in the motionless human test. On the opposite end of the arena, I set up two 1 m X 2.7 m liverpools (Figure 2a) for the training phase. Horses were kept in their personal stall (3.7 m X 3.7 m) when not in the testing arena.

Stimuli

Two liverpools (0.91 m X 2.74 m) were used for training for both experiments. For Experiment 1, one liverpool was striped vertically and the second was striped horizontally using duct tape. Both the vertical and horizontal stripes were 4.76 cm wide and were placed 15.24 cm apart on the liverpool.

For Experiment 2, one liverpool was colored blue and was filled with blue dyed water and sand. This made the water a semitransparent light brown color. The second liverpool was colored black and was filled with red/brown dyed water and sand. This made the water an opaque dark red/brown color. I dyed the water by putting Crayola Washable Finger-paint into a solo cup and mixing it with a small amount of water. The mixture from the cup was then poured into the water in the liverpool and mixed with a spoon until the desired color was reached.

For both experiments, I counterbalanced across horses the training they received (R-/R+ or R-) on which liverpool (vertically or horizontally taped liverpool, Experiment 1; blue or black liverpool with water, Experiment 2).

Sessions

1. *Schedule and Criteria*

Training consisted of training a horse to walk over a liverpool along the long axis. Horses and trainers started at the starting line, which was marked 2.44 m away from the liverpool (Figure 3). The target behavior was to walk across the liverpool without stopping or stepping off. Training occurred in four levels (Table 3). In Criteria 1, the trainer halted the horse before the liverpool and reinforced the behavior of one hoof entering the liverpoool at a time. The trainer delivered reinforcement for each hoof that entered the liverpool until all four hooves were on the liverpool. Horses were also halted for 5 s on the liverpool. The rescue Thoroughbred horses received reinforcement consistently after they halted for 5 s. However, the rest of the horses (all in Experiment 1 and some in Experiment 2) received reinforcement inconsistently due to trainer miscommunication. In Criteria 2, the trainer halted the horse before the liverpool and delivered reinforcement when all four hooves were on the liverpool. The horse then had to be halted on the liverpool for 5 s. As in Criteria 1, the rescue Thoroughbred horses received reinforcement consistently (once) after they halted for 5 s. The rest of the horses received reinforcement inconsistently due to trainer miscommunication. In Criteria 3, the trainer delivered reinforcement after all four hooves were on the liverpool and the horse had halted on the liverpool for 5 s. In Criteria 4, the trainer walked the horse through the liverpool without halting before or on the liverpool and delivered reinforcement once the horse crossed the liverpool and had all four hooves were off the liverpool. In Criteria 1-3, during the 5 s halt on the liverpool, horses had to remain straight and still for the 5 s when halted. If they mugged when halted on the liverpool, the trainer restarted the 5 s count. When 5 s elapsed some horses were reinforced. Others (Aegis, Chance, Grace, Murphy, Story, Cruz, Kate Spade, Riley, and. Squirt) were not reinforced in R-/R+ due to miscommunication between trainers. During the entire training session, if the horse bolted (see Table 4 for behavioral definitions), turned their hindquarters so that they were no longer parallel with the liverpool, or stepped sideways off the liverpool, they were brought back to the starting point and did not receive reinforcement (Figure 3). Regardless of what criteria a horse reached on any given day, they would restart at Criteria 1 the next day. This was to determine whether the horse reached a higher criteria with R- or R-/R+; that is, would the horse reach the same, higher or lower level with the other type of reinforcement the next day.

Training consisted of 10 min sessions per day for six days. The days alternated between R+/R- and R-. All horses started day one with R-/R+. Two trainers trained horses with R- and two trainers trained horses with R-/R+. Two trainers were graduate students from the Applied Animal and Welfare lab at Virginia Tech. They were both trained on proper timing and reinforcement delivery prior to this experiment through working with their own horses and teaching equine behavior and training classes. The two other trainers were undergraduate students. They were in a class prior to the study that taught them about proper timing and reinforcement delivery when training horses. Each horse was assigned one R- trainer and one R-/R+ trainer; the trainers assigned to each horse remained consistent throughout the study.

1. Combination Reinforcement (R-/R+)

In this condition, horses received both R+ and R- for correct responses. R+ consisted of a halved mini carrot; R- consisted of removal of halter and lead pressure.

The trainer cued the horse to move forward by saying “step up” and applying pressure to the halter. As soon as the horse moved forward the trainer said, “good,” and immediately released the pressure and delivered half of a mini carrot. The trainer cued the horse to halt by saying “whoa” and applying pressure. Once the horse halted the trainer released pressure said, “good”, and delivered half of a mini carrot.

Occasionally the horse engaged in undesirable behavior. If the horse pawed at the liverpool, slight pressure was applied to the halter until the behavior ended, then pressure was released, and no carrot was given. This was to prevent the liverpool from tearing. If the horse tried to mug, pressure was applied to the halter to move the head to forward position and once the horse’s head was in a forward position pressure was released.

However, Experiment 2 differed slightly because three horses, Pepino, Manny, and Roach stopped eating carrots. On the first day of R-/R+ training, Pepino ate the carrots until minute 8 of 10 and then stopped, and he would not eat grass the trainer offered. The trainer finished the session delivering scratching on the neck for about 3 s as a consequence. Roach ate a couple of carrots and then stopped. The trainer delivered grass clipping for the remainder of his session, which he consumed. Manny also ate a couple of carrots and then stopped. The trainer delivered grain for the remainder of his session, which he consumed. For the remaining combination training sessions, these three horses received grain as reinforcement (5 g, Proelite senior), which they all consumed.

2. Negative Reinforcement (R-).

This condition was conducted exactly like the R-/R+ condition except no carrot was delivered and I used different cues to ask the horse to move forward to help the horse discriminate conditions. In this condition, halter and lead pressure was removed contingent on the horse exhibiting the target behavior. The trainer cued the horse to move forward by saying “walk on” and applying pressure to the halter. As soon as the horse moved forward halter and lead pressure was removed. The trainer cued the horse to halt by saying “whoa” and applying pressure. As soon as the horse halted, halter and lead pressure were immediately released.

Measures

All training sessions were video recorded using a video camera and a tripod. I coded behaviors from video using Behavioral Observation Research Interactive Software (BORIS; Friard & Gamba, 2016). For Experiment 1 and 2, I coded frequency or duration of undesirable behaviors (frequency: horse stepping back, stepping to the side, pawing, biting, snorting; duration: head tossing, mugging, and for geldings, penis dropping) (Table 4). I coded these during time periods 0-1 min, 3-4 min, 6-7 min, and 9-10 min of each training session for days 2 and 3 in each condition.

Salivary Cortisol

Prior to the start of the experiment, I desensitized the horses to having a swab in their mouth for saliva collection. For the desensitization process I used Bounty paper towels and Mott’s Apple Flavored Applesauce. I rolled up paper towels until they reached the width of the SalivaBio Swabs. The rolled-up paper towels were placed into the jar of applesauce and rolled around until the applesauce was visible on the paper towel. Once the paper towel had applesauce on it, I placed it into the side of the horse’s mouth for 30 s. Once 30 s elapsed, the process was repeated on the same side of the horse’s mouth for another 30 s. After the second 30 s elapsed, I moved to the other side of the horse and placed the paper towel covered in applesauce in the horse's mouth for 30 s. On the fourth and final time, the paper towel was placed into the horse's mouth without applesauce. In Experiment 1, horses were desensitized three weeks before the experiment began. In Experiment 2, horses were desensitized two weeks before the experiment began.

Samples

I collected saliva samples using Salimetrics SalivaBio Children's Swab. I collected samples immediately before the horse's training session started (pre-training sample), and 30 min after their training session ended (post-training sample). In both cases, horses had a 30 min washout period from all types of feed and hay before saliva collection to reduce any contamination in the sample.

To collect the saliva sample, the research assistant held the lead rope attached to the horse's halter while the trainer collected the sample. The trainer wore gloves and placed the SalivaBio swab into the side of the horse's mouth for 30 s. During the 30 s, the trainer moved the swab underneath and on the side of the tongue. Once 30 s elapsed, the trainer cut off the dry portion of the swab and placed the swab portion that was in the horse's mouth into a 6 mL syringe. Using the syringe, the trainer expunged the saliva into a storage tube. If 1 mL of saliva was not collected, I repeated the process by putting the saliva swab back into the horse's mouth until I obtained 1 mL of saliva. The collection-expunging process was repeated until 1 mL of saliva was obtained. The storage tube was placed immediately into a cooler with ice. Once all samples were collected that day, samples were stored in a freezer until they shipped frozen overnight on dry ice to Salimetrics, seven weeks after data collection.

Samples were assayed at the Salimetrics' SalivaLab using the Salimetrics Salivary Cortisol Assay Kit (Cat. No. 1-3002). Samples were thawed to room temperature, vortexed, and then centrifuged for 15 minutes at approximately 3,000 RPM (1,500 x g) immediately before performing the assay. Samples were tested for salivary cortisol using a high sensitivity enzyme immunoassay (Cat. No. 1-3002). Sample test volume was 25 µl of saliva per determination. The assay has a lower limit of sensitivity of 0.007 µg/dL, a standard curve range from 0.012-3.0 µg/dL, and an average intra-assay coefficient of variation of 4.60%, and an average inter-assay coefficient of variation 6.00%, which meets the manufacturers' criteria for accuracy and repeatability in Salivary Bioscience, and exceeds the applicable NIH guidelines for Enhancing Reproducibility through Rigor and Transparency. I calculated the mean cortisol difference of pre- and post-training samples for each training condition for Experiments 1 and 2, and the mean cortisol level for each training condition for Experiment 2.

Motionless Human Test

Setting

Motionless human tests were conducted in the round pen before (pre-training) and after (post-training) each training session. In the round pen, I drew a 1 m radius circle in the middle by dragging a foot in the sand (Figure 4a). The person that served as the experimenter (the motionless human) was the trainer that was training the horse for their session that day.

The eight rehomed Thoroughbreds were tested at Virginia Tech's Smithfield Horse Center. I used a round pen (13.41 m diameter) of fourteen 3.05 m panels and a 1.83 m entrance gate (Figure 2b). I used a covered training barn to place horses in stalls (3.7 m X 3.7 m) when not being tested.

Sessions

The trainer stood in the middle of the 1 m radius circle in the round pen (Figure 4a). The assistant started the video camera and then walked the horse into the round pen. Once the horse was all the way through the entrance gate, the assistant unclipped the lead rope from the halter and started the timer and closed the entrance gate. Each test was 5 min, during which time the trainer did not make direct eye contact with the horse. They looked towards the ground or at the horse's hooves keeping the horse in their peripheral vision. If the horse came within arms' length of the trainer, the trainer petted and praised the horse. Petting consisted of scratching or rubbing the horses' body part that was in proximity to the trainer. Praising consisted of the trainer talking to the horse. Once the 5 min session ended, the assistant led the horse out of the round pen. The horse was then given a 2 min rest period during which the assistant or trainer held them by the lead rope until they were ready to have their pre-training cortisol sample collected. This process was repeated for their post training motionless test after the post training cortisol sample was taken.

Measures

I video recorded all sessions and coded all videos for time spent in proximity. I coded how long the horse was in proximity to the trainer and how long they interacted with the trainer. When the horse placed one hoof inside or on the 1 m radius circle, they were considered within proximity of the trainer (Sankey, Henry, et al., 2010; Sankey, Richard-Yris, Henry, et al., 2010; Sankey, Richard-Yris, Leroy, et al., 2010).

For Experiment 1 and some videos of Experiment 2, I hand coded the videos, due to technical issues with the SD cards. I did this by writing down when the horse started to be in proximity to the trainer, when proximity ended, and calculated time spent in proximity. For some videos of Experiment 2, I coded using Behavioral Observation Research Interactive Software (BORIS; Friard & Gamba, 2016). I coded when the horse was in proximity to the human.

Preference Test

After all training was completed (Day 7) I conducted a preference test for the two liverpools used in training to evaluate whether horses preferred the liverpool associated with R-/R+ or the liverpool associated with R-.

Setting

I used the same round pen as was used in the motionless human test (16.76 m diameter) and placed both liverpools inside the round pen (Figure 4b). Liverpools were placed 2.44 m apart from each other in the middle of the round pen. A 1 m perimeter was drawn by the experimenter using their hand or foot into the sand around each liverpool.

The eight rehomed Thoroughbred horses were tested at Virginia Tech's Smithfield Horse Center. I used a round pen (13.41 m diameter) of fourteen 3.05 m panels and a 1.83 m entrance gate (Figure 2b). I used a covered training barn to place horses in stalls (3.7 m X 3.7 m) when not being tested.

Stimuli

Horses in Experiment 1 were tested with the vertically and horizontally striped liverpools; horses in Experiment 2 were tested with the blue and black liverpools.

Sessions

Each horse completed two preference tests, with the side on which each liverpool was presented alternated across the tests to try to account for any side biases. Thus, half the horses had the combination stimulus on the right in their first preference test, and half the horses had the combination stimulus on the left in their first preference test; in the second preference test, the combination stimulus was on the opposite side for each horse. Once all horses completed the first test, they were put away in stalls while the experimenter switched the liverpools. Once the liverpools were switched, horses completed the second test. Horses performed the second test approximately 30 to 70 min after completing the first test.

Horses were led into the round pen by the trainer or research assistant and were released when they were equidistant between the two liverpools and facing towards the liverpools (Figure 4b). Sessions lasted for 5 min. Once the 5 min elapsed, the horse was led out of the round pen and placed into a stall.

Measures

I measured the horse's initial choice, duration they were in proximity to each liverpool, and the duration they interacted with each liverpool. A choice was defined as one hoof crossing into or on the 1 m perimeter surrounding one of the liverpools. Proximity was defined as the horse having at least one hoof stayed in or on the 1 m perimeter. Interaction was defined as the horse touching the liverpool with a hoof.

I video recorded all sessions. Two 5 min videos were recorded for each horse for each test to get angles on both liverpools. For Experiment 1 and some videos of Experiment 2, I hand coded the videos due to technical issues with the SD card. I did this by writing down when the horse started to be in proximity to or interact with the liverpool and when proximity and interaction ended. I also recorded their initial choice. For some videos of Experiment 2, I coded from video using Behavioral Observation Research Interactive Software (BORIS; Friard & Gamba, 2016).

Independent Observer Agreement

A second, independent observer double-coded 20% of videos for interobserver agreement (IOA). I calculated IOA as a percentage of time as long as they were more than 3 s different. If less than 3 s different, they were counted as an agreement. Mean IOA for horses in proximity to the trainer for the motionless human test was 90% and the mean IOA for horses in proximity to the liverpool in the preference test was 91%.

RESULTS AND DISCUSSION

Training

Ten horses completed training during Experiment 1 and 16 horses completed training in Experiment 2. In Experiment 1, I had to drop the first day in each training condition for Aegis, Bentley, Chance, Gibson, and Grace because I did not have clear criteria steps. This created problems with the trainers training equally. The remaining horses (Java, Murphy, Pumba, Story, and Dragonfly) only trained for two days in each condition because one trainer had a last-minute emergency and could not be there on the first day of data collection.

To see how successful each training condition was, I looked at criteria reached in the first and last day of each condition. In Experiment 1, I had to drop the first day in each training condition for Aegis, Bentley, Chance, Gibson, and Grace; thus, I only analyzed the criteria reached for Day 3. All five horses reached Criteria 4 in both conditions on Day 3.The remaining horses in Experiment 1 received training for two days in each condition. I looked at criteria reached in Day 1 and Day 2 in each condition (Table 5a and 5b). For both Day 1 and Day 2, all five horses reached Criteria 4 in R- and zero horses reached Criteria 4 in R-/R+. A Chi-Square test indicated that the proportions of criteria reached in Day 1 and Day 2 were statistically significant, Likelihood Ratio for both Day 1 and Day 2, X^2 (3, $N = 10$) = 13.86, $p < .05$ and Pearson Chi- Square for both Day 1 and Day 2, X^2 (3, $N = 10$) = 10, $p < .05$. Since horses had multiple rows of data, I blocked the horse's name by performing a Cochran-Mantel-Haenszel Test, to further ensure that the proportions were statistically significant from expected proportions. The Cochran-Mantel-Haenszel Test indicated that the proportions were not statistically significant from expected proportions for Day 1, $X^2_{MH} = 5.00$, 3 d.f., $p > .05$, and Day 2, $X^2_{MH} = 5.00$, 3 d.f., $p > .05$.

For Experiment 2, on Day 1 only one horse reached Criteria 4 in both conditions. In R-/R+, 13 horses did not reach Criteria 1, two horses reached Criteria 1, and one horse reached Criteria 3 (Table 5c). In R-, 12 horses did not reach Criteria 1, two horses reached Criteria 1, and one horse reached Criteria 2. The contingency table for Day 3 (Table 5d) indicated that proportions of criteria levels was similar between conditions. A Chi-Square test indicated that the proportions of criteria reached in Day 1 were not statistically significant, Likelihood Ratio, X^2 (4, $N = 33$) = 2.78, $p > .05$ and Pearson Chi- Square, X^2 (4, $N = 33$) = 2.01, $p > .05$ or Day 3, Likelihood Ratio, X^2 (3, $N = 31$) = 7.80, $p > .05$ and Pearson Chi- Square, X^2 (3, $N = 31$) = 5.86, $p > .05$.

When looking at criteria reached for Experiments 1 and 2, the proportions of each criteria in each condition for Day 1 and Day 3 were not statistically different. In Experiment 1, the five horses that were trained for all three days all reached Criteria 4 in both conditions on Day 3. The other five horses in Experiment 1, on Day 1 and Day 2 reached Criteria 4 in R-, but zero horses reached Criteria 4 in R-/R+. These results may be due to the addition of R+. The addition of R+ caused horses to mug for longer durations and therefore spent less time to achieve all criteria levels. In Experiment 2, seven horses reached Criteria 4 in R-/R+ and eight horses reached Criteria 4 in R- on Day 3. However, on Day 3, in R-/R+ seven horses did not reach Criteria 1 and in R- four horses did not reach Criteria 1. The reason why all horses in Experiment 1 reached Criteria 4 in each condition could be due to the stimuli used. In Experiment 1, the liverpools were taped and not filled with water, whereas in Experiment 2 the liverpools had colored water in them. Looking at the behavioral data, it suggests that the horses in Experiment 2 viewed the liverpool as a more aversive stimulus. Using a Wilcoxon signed-rank test, I found that the total occurrences of stepping back regardless of training condition, horses in Experiment 2 (*Mdn*= 5, *IQR*= 1-8) stepped back more frequently than horses in Experiment 1 (*Mdn*=0, *IQR*= 0-0.25), Z = 2.72, $p < .05$.

When evaluating other behaviors that horses exhibited during training, stepping to the side, stepping back, pawing, dropping their penis, mugging, and head tossing were seen in both training conditions in Experiment 1. However, biting was only seen in R- and snorting was only seen in R-/R+. I used the mean frequency or duration across all days that I had data for to calculate median frequencies or durations for each horse. Thus, each horse was represented by one datum for each behavior in each training type in my analysis except that only five horses trained for two days in each condition. I analyzed the results from Day 3 for Aegis, Chance, Grace, Murphy, Story. I used Wilcoxon signed-rank tests to evaluate the difference in frequency of point behaviors and duration of state behaviors between R-/R+ and R- training. I found that horses stepped back more frequently in R-/R+ (*Mdn* = 3, *IQR* = 0-3) than R- (*Mdn* = 0, *IQR* =0-1*),* Z = -2.82, $p < .05$ (Figure 5a) and that horses engaged in mugging for longer periods of time in R-/R+ (*Mdn* = 6 s, *IQR* = 2-12.498) than R- (*Mdn* = 1.50 s, *IQR* = 0-3.76), Z = -2.41, $p < .05$ (Figure 5b). There were no other significant differences in behaviors between training types.

In Experiment 2, similar to Experiment 1, stepping to the side, stepping back, dropping their penis, head tossing, and mugging were seen in both training conditions. Biting was only seen in R- and snorting was not seen in either condition. I analyzed results from Day 2 combined with Day 3 in each condition by comparing the total duration or total occurrence of each behavior that was seen during each training type. Using a Wilcoxon signed-rank test, I found that horses engaged in mugging for a longer period of time in R-/R+ (*Mdn* = 1.62 s, *IQR* = 0.50-6.81) than R- (*Mdn* = 0 s, *IQR* = 0-0), Z = -3.81, $p < .05$ (Figure 6a). There were no other significant differences in behaviors between training types (Figure 6b).

For Experiment 1, horses stepped back more frequently in R-/R+ than R-. With a greater frequency of stepping back, horses also had a greater variance of stepping back in R-/R+ than R-. The greater frequency and variance of stepping to back could be due to horses trying to approach the trainer to access carrots. After receiving a carrot, the horse's motivation might have changed and they engaged in new behaviors to access the carrot. Horses might have also been confused with what the trainer was asking them to do with the unfamiliar delivery of carrot as reinforcement and therefore took more steps back rather than stepping forward, unlike in R-.

In both experiments horses engaged in mugging for significantly longer durations in R-/R+ training than R-. This is likely due to the addition of the carrot as an appetitive stimulus and the horse's learning history. Most likely, these horses have received little R+ and have not been trained to not mug when food is present. Kurland (2001) recommends explicitly training horses to not mug the trainer when R+ is used. I did not train these horses to be polite around food (i.e., not mug) prior to beginning the experiment. That mugging occurred for longer durations in the R-/R+ condition is thus not surprising. My data indicates that food delivery in general can bring about increased oral behaviors and parallels prior research (Hockenhull & Creighton, 2010). The results from Fox et al. (2012), however, suggest that this mugging behavior can be solved through clear, appropriate R+ contingencies.

Overall, between the criteria reached data and the behavioral observation data, I found only a few differences in behavior between training types. There are a few reasons as to why only slight differences were seen. First, horses received a total of 30 min (10 min sessions for three days) in each condition. With increased duration of training sessions or more training sessions, the horses would have had longer exposure to the different contingencies and more behavioral differences might have emerged. Additionally, using different magnitudes of reinforcement could have produced more behavioral differences. The magnitude R+ used in R-/R+ might not have been large enough to be salient. In this study one half of one mini carrot was delivered for correct responses. Increasing the magnitude of reinforcement to a larger magnitude of carrot or even to a different type of food, such as grain, might produce larger differences in behavioral measurements and criteria reached between the two training types. In fact, two rehomed Thoroughbreds, Manny and Roach, would not eat the carrots and they only consumed grain. Once I switched to grain, the horses were more willing and engaged in the training. R- was also very mild within this study and is unlikely to elicit much aggression. The pressure on the halter is a very mild and very standard aversive stimulus for managing horse behavior. Using a more aversive stimulus could likely produce more aggression than I observed, similar to the results of Azrin et al., (1964).

Additionally, the use of a within-subjects design imposed some limitations when it comes to analyzing the training data. While the use of a within-subjects design can be useful to reduce individual differences, especially in acquisition tasks, there is a greater probability of carry over effects between conditions that might obscure differences between the two training types. That is, what the horses learned on Day 1, regardless of training type, likely impacted their performance on the next day, making assessing possible differences in training efficacy more challenging.

Cortisol

I collected pre- and post- training salivary cortisol samples from all horses. All samples were analyzed in duplicate, except for Shenandoah's R- Day 3 pre-sample and Murphy's R- Day 1 post-sample. I used the mean of the duplicate samples in my analysis or the value provided by the analysis completed for the two samples not analyzed in duplicate. I used a Wilcoxon-signed rank test to test whether the difference in cortisol levels (pre- minus post-training) differed between R-/R+ and R- training. I found there was no statistical difference between R-/R+ (*Mdn* = -0.007 , *IQR* =-0.03-0.02) and R- (*Mdn* = -0.005, *IQR* = -0.02-0.01) in Experiment 1, Z = 0.44, $p > .05$, and no statistical difference between R-/R+ (*Mdn* = -0.008, *IQR* = -0.04-0.01) and R- (*Mdn* = -.007, *IQR* = -0.03-0.008) in Experiment 2, Z = -0.05, $p > .05$ (Figure 7). Using a Brown-Forsythe to discover if the variances of cortisol difference levels differed between R-/R+ and R-, I found that there was greater variance in the R-/R+ training $F(1,47) = 4.9491$, $p < .05$ in Experiment 1 but that there was no statistically significant difference between the variance of R-/R+ versus R- in Experiment 2 $F(1,97) = 0.4952$, $p > .05$.

I used a multiway ANOVA to test whether cortisol differences might have been impacted by the following factors: type of training, number of times they stepped back, pawed, snorted, bit, and stepped to the side, and the duration of mugging, penis dropping, and head tossing. For Experiment 1, I ran an ANOVA for Days 2 and 3 in each condition. The factors listed above had no significant effect on cortisol difference ($F(9,20) = 0.9772$, $p > .05$, $R^2 = 0.31$, $R^2_{Adjusted} = -0.01$). For Experiment 2, I looked at Days 1, 2 and 3, in each condition. I analyzed Day 1 in each condition separately from Day 2 and 3 since Day 1 training was dropped for some of the horses in Experiment 1. None of the factors had a significant effect on cortisol difference for Day 1, ($F(9,20) = 0.4534$, $p > .05$, $R^2 = 0.12$, $R^2_{Adjusted} = -0.20$), or for Days 2 and 3, ($F(8,50) = 0.6864$, $p > 0.05$, $R^2 = 0.10$, $R^2_{Adjusted} = -0.05$).

For Experiment 1, I ran a two-way ANOVA to test how training condition impacted pre- and post-training cortisol levels for Day 1 and Day 2. There was no statistical effect of pre- and post-training cortisol levels (F (1.73,29.39) = 0.60, p < .05). There was no statistical effect for reinforcement condition (F (1,17) = 3.57, p < .05. There was also no statistical effect between the interaction of pre and post-training levels and reinforcement condition ($F(3,51) = 0.18$, $p < .05$). For Experiment 2, I ran a two-way ANOVA to test how training condition impacted pre and post-training cortisol levels for Day 1, 2, and 3. There was no statistical effect of pre- and post-training cortisol levels ($F(2.05, 59.52) = 2.10$, $p < .05$). There was no statistical effect for reinforcement condition ($F(1,29) = 1.77$, $p < .05$. There was also no statistical effect between the interaction of pre- and post-training levels and reinforcement condition ($F(5,145) = 0.31$, $p < .05$).

I found no statistical differences in changes in cortisol between R- and R-/R+. This suggests that neither training type induced more stress or arousal than the other. My results parallel previous work in which dogs trained with a shock collar or with food as a reinforcer showed similar levels of urinary and salivary cortisol (Cooper et al., 2014). Additionally, while both aversive and appetitive stimuli can increase cortisol levels (Merali et al., 1998) my results did not find that delivery of food or the removal of the aversive stimulus increased salivary cortisol levels. However, I used relatively low intensity R- and R+ in my study. If I had utilized larger magnitudes of aversive or appetitive stimuli, it is possible that I would have seen an increase in cortisol levels in one or both conditions. For example, when (Beerda et al., 1998) applied an aversive stimulus (shock and loud noise) to dogs, cortisol levels increased.

Because cortisol levels do not always seem to reflect potential stressors, behavioral observations can be useful to provide additional information about the state of the animal (Harewood & McGowan, 2005). In this study I saw few differences between training types; besides increased mugging in both experiments and increased stepping back in Experiment 1 with R-/R+ training, total occurrence and duration of behaviors did not differ between training conditions. However, it is worth noting that one behavior was exclusively seen in one training condition. Biting was seen only in R- but was not seen in R-/R+. The fact that I saw no cortisol differences parallels my findings of no behavioral differences.

Oxytocin is released in response to a number of stressors in a way to decrease stress hormones (Carter & Lightman, 1987). Oxytocin levels could have affected cortisol levels in my study, however it is unlikely due to previous research. Previous research was conducted looking at cortisol and oxytocin levels of horses during pre- and post-grooming sessions (Lansade et al., 2018). There was no statistical difference between cortisol and oxytocin levels during pre- and post-grooming sessions. My data indicated no statistical differences in changes in cortisol between R- and R-/R+. My data aligns with previous research and indicates that although we did not collect oxytocin levels, they unlikely affected cortisol levels.

Motionless Human Test

Using the difference in time spent with in proximity to the human in the pre- and post-tests, I categorized the data into three categories: negative, if the time horses spent in proximity to the human decreased from pre- to post-training; positive, if the time horses spent in proximity to the human increased from pre- to post-training; or no change, if the time horses spent in proximity to the human did not differ (within 1 sec) from pre- to post-training. I made a contingency table for Experiment 1 and Experiment 2 to look at how frequency these categorical variables occurred in each training condition (Table 6).

The contingency table for the difference in time spent in proximity in Experiment 1 (Table 6a) indicated that more horses in R-/R+ showed a greater proportion of negative change or positive change than seen for R-. That is, more horses in R-/R+ spent more time with the R+/R- trainer post-training than pre-training (n = 9) than when tested with the R- trainer (n = 5), but also more horses in R-/R+ spent less time with the R+/R- post-training than pre-training(n =

12) than when tested with the R- trainer (n = 10); whereas more horses showed no change in time spent with the trainer pre- to post-training when tested with the R- trainer (n = 10). A Chi-Square test indicated that the proportions were statistically significant, Likelihood Ratio, X^2 (2, N = 48) = 7.1, $p < .05$ and Pearson Chi- Square, X^2 (2, N = 48) = 6.6, $p < .05$. Since horses had multiple rows of data, I blocked the horse's name by performing a Cochran-Mantel-Haenszel Test, to further ensure that the proportions were statistically significant from expected proportions. The Cochran-Mantel-Haenszel Test indicated that the proportions were statistically significant from expected proportions, X^2_{MH} = 6.62, 2 d.f., $p < .05$.

The contingency table for the difference in time spent in proximity in Experiment 2 (Table 6b) indicates that for R-/R+ there was a greater proportion of horses that showed a positive change than seen for R-. That is, more horses in R-/R+ spent more time with the R+/R- trainer post-training than pre-training (n = 20) than when horses were tested with the R- trainer (n = 8). When tested with R- trainer, more horses showed no change in time spent with the trainer from pre- to post-training (n = 25) than when tested with the R-/R+ trainer (n = 15). That is, more horses spent the same time with the trainer pre- and post-training for R-. A Chi-square test indicated that the proportions were statistically significant. The Likelihood ratio, X^2 (2, N =98) = 7.8, $p < .05$ and Pearson Chi-square, X^2 (2, N = 98) =7.6, $p < .05$, indicate that the proportions are statistically significant from expected proportions. I then ran a Cochran-Mantel-Haenszel test. It ensured that there was a statistically significant from expected proportions, X^2_{MH} = 8.92, 2 d.f., $p < .05$.

In both experiments, when tested with the R-, more horses tended to show no change in time spent with the trainer. This finding parallels that of Sankey, Richard-Yris, Leroy, et al., (2010) in that horses trained with no reinforcement compared to R+, spent less time with trainers. In neither experiment did more horses show more time spent with the R- trainer after training. However, this was expected as that trainer had been paired with aversive stimuli and not appetitive stimuli. Along these same lines, prior research found that horses decreased their time spent with a trainer not associated with R+ (Sankey, Richard-Yris, Henry, et al., 2010; Sankey, Richard-Yris, Leroy, et al., 2010). It is possible that I did not see a decrease in time spent with the R- trainer because horses alternated between R-/R+ and R- and the positive association produced by R+ could have generalized to the R- trainer similar to what was observed by (Sankey, Richard-Yris, Leroy, et al., 2010). Additionally, the magnitude of reinforcement might have also played a part on horses showing no change instead of less time. In the current study, I applied pressure to the halter and released it contingently training. This is a mild and typical aversive stimulus used in daily horse operations. If a more intense aversive stimulus was used in the R- training such as using a crop or putting a chain over the horses' nose and then applying and releasing pressure, horses likely would have spent less time with R- trainer than showing no change.

In both Experiments 1 and 2, a greater proportion of horses spent more time after training than before training when tested with the R-/R+ trainer. This would have been predicted given that the trainer was being paired through classical conditioning with an appetitive stimulus (Baragli et al., 2015; Starling et al., 2016; York et al., 2017). My results parallel what Sankey, Henry, et al., (2010) found when they performed their motionless human test with horses trained with food or grooming: horses spent more time with their trainer in the food reward group than the grooming reward group. In another study Sankey, Richard-Yris, Leroy, et al., (2010) also found that in their short-term test, horses spent more time in proximity to their trainer after training with R+. Their data and my data suggest that using R+ increases affiliative behavior in horses towards their trainers. My data also suggests that the more time in proximity to their trainer and an increase in affiliative behavior after R+ could be an independent effect of classical conditioning of the trainer with food.

It is likely that my results parallel those of Sankey, Richard-Yris, Leroy, et al., (2010), because horses associated their R-/R+ as a positive experience with their trainer. Because I used a low intensity aversive stimulus in R-, and one that these horses were used to experiencing, it is likely that the R+ part of the training was salient to them, despite also receiving R-. This is supported by the fact that when horses were tested with trainer associated with R-, they showed no change in time spent with their trainer pre- versus post- training. These horses were probably use to R-, therefore they could have associated their trainer as a neutral stimulus rather than a negative stimulus despite the trainer being associated with R-.

Nevertheless, some horses in Experiment 1 spent less time with the R-/R+ after training in Experiment 1. It is possible that extinction could have occurred and produced a decrease in time spent in the post-training test. During training they had been receiving carrots from the trainer, but in this test no carrots were given. The contrast between reinforcement and extinction could have produced a reduction in time engaged for some horses (McCall & Burgin, 2002).

While my results parallel prior studies using R+, (Sankey, Henry, et al., 2010; Sankey, Richard-Yris, Leroy, et al., (2010), these studies did not compare R-/R+or R-. My study adds how R- impacts the human-animal bond and also demonstrated that even when R+ is used in conjunction with R-, it can increase the time the horse spends with the human.

Preference Test

In Experiment 1, Test 1, three horses chose R-/R+ first, four horses chose R- first, and three horses did not make a choice. In Test 2, five horses chose R-/R+ first, two horses chose R-first, three horses did not make a choice. Only one horse, Chance, picked R-/R+ for both preference tests. No horses chose R- in both tests. In Experiment 2, Test 1, six horses chose R-/R+ first , five horses chose R- first, and four horses did not make a choice. In Test 2, five horses chose R-/R+ first, two horses chose R- first, and eight horses did not make a choice. Two horses, Roach and Shenandoah, chose R-/R+ for both tests. Two horses, Manny and Hijack, chose R- for both tests. I performed a binomial test on the horses' first choice in Test 1 and Test 2. In none of these (Experiments 1 and 2, Tests 1 and 2) did horses choose either the R-/R+ or R- liverpool at above chance levels (smallest $p = .11$ for choosing the negative stimulus in Test 1 of Experiment 1).

I used an ANOVA to look at time spent in proximity to either liverpool with the factors of horse type (riding herd or rescue Thoroughbreds), stimulation type (R- +R+ or R-), liverpool type (horizontal or vertical, back or blue) and test number (1 or 2). I found that stimulation type, liverpool type and type of horse had no significant effect on the time the horse spent in proximity to either liverpool in Experiment 1, ($F\,(4,35) = 1.08$, $p > .05$, $R^2 = 0.11$, $R^2_{Adjusted} = 0.01$) or in Experiment 2, ($F\,(4,55) = 1.37$, $p > .05$, $R^2 = 0.09$, $R^2_{Adjusted} = 0.02$). The overall model did not explain the variance in the response (preference), but we did find that the horse's preference mattered in explaining the response ($F\,(1,55) = 4.51$, $p < .05$).

In general, the findings from the preference test showed no difference between R-/R+ and R- in either experiment. For the horse's first choice, there was no difference between choosing the R- or R-/R+ stimulus. They also show that R- and R-/R+ did not correlate with how long they spent in proximity to the stimulus.

This suggests that horses did not prefer one type of training over the other. However, Sankey, Henry, et al. (2010), Sankey, Richard-Yris, Henry, et al. (2010) , and Sankey, Richard-

Yris, Leroy, et al. (2010) found that R+ training produces increased allocation of time to trainers associated with R+, this seems unlikely. The lack of preference might be due the horses not discriminating between the stimuli or associating one stimulus with the specific training. The liverpools themselves might not have been sufficiently salient and the horse learned to associate the trainer but not the liverpool with the training type. Longer training or larger magnitudes of aversive or appetitive stimuli might have produced a clearer preference in horses.

Additionally, Mejdell et al. (2019) found that multiple factors play a role when the horse made a choice between blankets. The factors included wind, precipitation, and air temperature. Although my analysis included, horse type, liverpool type, stimulation type, and test number, it is possible that I did not measure a factor that influenced the horse's preference. In Experiment 1, six horses chose the same side in Test 1 and Test 2 regardless of the type of stimulus. In Experiment 2, three horses chose the same side in Test 1 and Test 2 regardless of the type of stimulus. This data is evidence that for some horses a side bias effect occurred.

Finally, not all horses made a choice in the preference test. For Test 1, 11 out of 15 horses made a choice and in Test 2, only seven out of 15 horses made a choice. The decrease in choices made from Test 1 to Test 2 was likely impacted by stimulus novelty. In Test 2, horses had already interacted with the liverpools in Test 1. Additionally, because I tested them in an arena with visual and auditory access to the environment, environmental stimuli could have caused the horses to be more interested in those rather than engaging in the preference test.

Interestingly, the eight horses that did not make a choice in Experiment 2 Test 2 were all part of the Virginia Tech riding herd. This could be because the environment was less novel to them because they are ridden in that arena for lessons and they explored less or because, in that environment, exploration and interaction with stimuli could have been punished.

CONCLUSIONS AND FUTURE DIRECTIONS

The use of R+ has been increasing in the equine industry. However, since many horses that might have started in R- training might now have R+ layered on, understanding the effects of combination training (R-/R+) can help us understand the impacts on horses as they transition into new training techniques.

My results add to the knowledge of how this transition period may affect horses compared to how they have been trained for many years. In general, I found few differences between training horses with R-/R+ or R-. Horses reached similar training criteria, other than mugging did not reliably show differences in other observed behaviors, the changes in cortisol levels did not differ across training types, and I did not find a preference for either stimulus associated with a specific type of training. My findings of similar training criteria being reached and no significant differences in observed behaviors, besides mugging, align with the findings from Heleski et al. (2008). When comparing R- to R-/R+ they also found there was no statistical difference of horses reaching calmness criterion in R-/R+ or R-. They found the addition of R+ did not enhance the learning task. My results also parallel those that Olczak et al. (2018) found when using an appetitive stimulus. Their results showed little difference in pre- and post- testing results, suggesting that the training and testing might not have been stressful.

It is possible that for horses that have had a substantial reinforcement history with R-, the addition of R+ does not impact their ability to learn, possible behavioral side-effect, or stress as measured by cortisol levels. It is possible that looking at differences in R-/R+ and R- might be more apparent in young horses that have not already learned how to effectively respond to pressure used in R-. As Sankey et al. (2011) suggested with their findings, naïve horses displayed more negative behaviors and were reluctant to human touch when approached by humans;

however, older horses showed more positive behaviors. Using a naïve horse rather than an older horse with a longer history of reinforcement might produce greater differences in behavioral responses when looking at R- vs. R-/R+.

The experimental parameters in my study might have influenced the lack of differential outcomes. First, the horses only had three 10-min sessions for each training type. The lack of observed differences could have been due inadequate contact with the contingencies and differences could have emerged with extended contact with R-/R+ and R-. Given that these horses had likely been trained largely if not completely with R- during their lives, 30 min of R-/R+ training might not have been enough to produce differential outcomes. Additionally, the stimuli used were also relatively low magnitude: I used half baby carrots and pressure on the halter. A larger positive reinforcer could have been more salient, although the increase in mugging behavior does indicate that the addition of the carrot was salient to some degree for the horses, and the level of aversive stimulation is such that the horses were likely very familiar to the horse. Looking at the effects of larger intensity stimuli could also produce greater differences in the behavioral and physiological measures I analyzed.

One of the largest issues that could impact the lack of differential outcomes is the possibility of carryover effects. That is, what a horse learned on Day 1, regardless of the type of training, carries over to their performance on the subsequent day. While, I used visually distinct stimuli (vertically- or horizontally-taped liverpool and a blue or black liverpool with colored water) in hopes that horses would discriminate between the two tasks, because the tasks were identical there were likely carryover effects. While one way to eliminate this is using group designs, this introduces individual differences as potential sources of error. Another solution is to use different tasks that are equally challenging, which can be hard to determine but might be a useful future direction. It is hard to determine if tasks are equally challenging based on horses individual differences.

Nevertheless, there were two areas where I did detect differences. The first area is mugging which was significantly higher in R-/R+ in both experiments. The mugging behavior suggests that the horse knew that the carrot was available. My results align with results those of Hockenhull and Creighton (2010) in which they found that roughly and gently searching clothing and hand licking was significantly associated with handfeeding. This mugging behavior could have been reduced if horses had been trained prior to not mug when food reinforcement was provided.

The second area where I detected a difference was during the motionless human test; my results showed that in both experiments a greater proportion of horses spent more time after training than before training with the trainer when tested with the R-/R+ trainer. My results parallel those that Sankey, Henry, et al. (2010); Sankey, Richard-Yris, Henry, et al. (2010); Sankey, Richard-Yris, Leroy, et al. (2010) found: horses trained with an appetitive stimulus spent more time in proximity to their trainer. My results further add to the possibility that incorporating R+ into training can improve the human-horse relationship by producing more affiliative behaviors from the horse to the human.

While my study did not show large differences in R-/R+ compared to R-, this area necessitates more research evaluating a range of parameters that could influence outcomes, such as task trained, length of training, and magnitude of reinforcement, as well as evaluating the effects of these training types in more horses, of different breeds, and different training backgrounds. For instance, in my study, regardless of training condition, the rehomed

Thoroughbreds had a higher pre- training cortisol level (*M*= 0.202, SD=0.19) than the riding herd (*M*= 0.07, SD= 0.05). This difference may be due to the fact that the Thoroughbreds are kept on pasture and rarely have to go into a stall. Where the riding herd stay in stalls for part of the day. Since the Thoroughbred horses were not used to being in a stall, this could be a reason as to why their pre-training cortisol levels were higher than the riding herd's levels. The Thoroughbred horses also started their training with R- where the other horses in Experiment 2 started with R-/R+. However, the breed differences cannot be disentangled from the training order. Thus, one thing that could be explored in the future is analyzing if breed differences played an effect on cortisol results.

Nevertheless, the differences seen in time spent with trainer do indicate that the addition of R+ to R- training, even over a short period of time, can influence the level of affiliative behaviors horses might show to the trainer. Although in this study there were few differences seen between R-/R+ and R-, adding R+ into R- training methods should still occur. For example, (Heleski et al., 2008) found few differences between R-/R+ and R-, but through their observations during the experiment they noted that the experimental environment was safer for horse and handler during R-/R+ than R-.

TABLES

a.

Name	Sex	Age	Breed*	First Condition
Aegis	G	21y	American Warmblood	Combination
Bentley	G	11y	Warmblood	Combination
Chance	G	18y	Thoroughbred	Combination
Gibson	G	8y	Welsh/Thoroughbred	Combination
Grace	M	7y	Appendix Quarter Horse	Combination
Java	M	14y	Paint	Combination
Murphy	G	15y	Welsh Pony	Combination
Pumba	G	18y	Westphalian	Combination
Story	G	13y	Thoroughbred	Combination
Dragonfly	G	15y	Oldenburg	Combination

b.

Name	Sex	Age	Breed*	First Condition
Artimus	M	7y	Thoroughbred	Combination
Commodore	G	21y	Thoroughbred	Combination
Cruz	G	15y	Thoroughbred	Combination
Del Fino	G	13y	Hanoverian	Combination
Kate Spade	M	7y	Trakehner	Combination
Mai Tai	M	8y	Trakehner	Combination
Riley	G	16y	Welsh Pony	Combination
River	G	22y	Quarter Horse	Combination
Squirt	G	18y	Paint	Combination
Kwil	G	15y	Dutch Warmblood	Combination

c.

Name	*Sex*	*Age*	*Breed*	First Condition
Bobkat	G		Thoroughbred	Negative
Pepino	G		Thoroughbred	Negative
Sondheim	G		Thoroughbred	Negative
Shenandoah	G		Thoroughbred	Negative
Hijack	G		Thoroughbred	Negative
Manny	G		Thoroughbred	Negative
Roach	G		Thoroughbred	Negative
Ranger	G		Thoroughbred	Negative

Table 1. Demographic data of the horses. Age is reported in years (y). Sex: Gelding (g) and mare (m). Horses' breeds are listed, but not all breeds were confirmed with registration papers. First condition states which training condition horses started in on their first day of training (R-/R+ or

R-). *a.* Horses in Experiment 1. *b.* Horses from the Virginia Tech hunt seat herd in Experiment 2. *c*. Rehomed Thoroughbreds from the Virginia Tech research herd in Experiment 2.

Task Number	Task
1	Human Motionless test (Pre-)
2	Rest (2 min)
3	Salivary cortisol sample collected (Pre-)
4	Training (10 min)
5	30 min washout period
6	Salivary cortisol sample collected (Post-)
7	Rest (2 min)
8	Motionless human test (Post-)
9	Horses placed back into their stalls and given hay

Table 2. Daily schedule that indicates the order in which motionless human tests, training, and cortisol samples were completed for Experiments 1 and 2.

a.

Task Number	Task
1	Walk up to the liverpool and halt
2	Move horse forward until one hoof is on the liverpool and halt
3	Move horse forward until two hooves are on the liverpool and halt
4	Move horse forward until three hooves are on the liverpool and halt
5	Move horse forward until four hooves are on the liverpool and halt
6	Let horse stand for five seconds with head forward or slightly away from experimenter
7	Walk off the liverpool until all four hooves are off and halt
8	Line up on the other side at the start line
9	Repeat the following steps twice in each direction

b.

Task Number	Task
1	Walk up to liverpool and halt
2	Move horse forward until four hooves are on the liverpool and halt
3	Let horse stand for five seconds with head forward or slightly away from experimenter
4	Walk off the liverpool until all four hooves are off and halt
5	Line up on the other side at the start line
6	Repeat the following steps twice in each direction

c.

Task Number	Task
1	Move horse forward until all four hooves are on the liverpool and halt
2	Let horse stand for five seconds with head forward or slightly away from experimenter
3	Walk off the liverpool until all four hooves are off and halt
4	Line up on the other side at the start line
5	Repeat the following steps once in each direction

d.

Task Number	Task
1	Walk all the way across the liverpool and halt once all four hooves are off the liverpool
2	Repeat until training time is up

Table 3. Training criteria. The horse had to meet each criterion before moving on to the next criteria in either R-/R+ or R-training. *a.* Steps the trainers followed to complete Criteria 1. *b.* Steps the trainers followed to complete Criteria 2. *c.* Steps the trainers followed to complete Criteria 3. *d.* Steps the trainers followed to complete Criteria 4.

Behavioral Definitions

Behavior	Definition
Bolting	The horse moved past the trainer without stopping from pressure,
Steps Back	Horse took a step back with front hooves any time during the training session
Steps to the Side	Horse stepped to the side with front hooves any time during the training session
Pawing	Horse lifts the front leg up then drags their hoof across the ground or liverpool
Biting	Horse shows teeth or bites toward trainer or lead rope
Drops Penis	Horse's penis drops outside of sheath (geldings only)
Head tossing	Horse moves its head up and down at a fast pace at least once
Mugging	Horse reaches for treat pouch or turns head and touches trainer, instead of standing with head forward or away from trainer
Snorting	Horse makes a short noise from the vibration of forceful air exiting its nostrils

Table 4. Behavioral definitions for R-/R+ and R- training.

5a: Experiment 1: Criteria Reached Day 1

		Criteria 0		Criteria 1		Criteria 2		Criteria 3		Criteria 4	
		n	%	n	%	n	%	n	%	n	%
Type of	R-/R+	1	10	2	20	2	20	0	0	0	0
Training	R-	0	0	0	0	0	0	0	0	5	50
Total:		1	10	2	20	2	20	0	0	5	50

5b: Experiment 1: Criteria Reached Day 2

		Criteria 0		Criteria 1		Criteria 2		Criteria 3		Criteria 4	
		n	%	n	%	n	%	n	%	n	%
Type of	R-/R+	0	0	3	30	1	10	1	10	0	0
Training	R-	0	0	0	0	0	0	0	0	5	50
Total:		0	0	3	30	1	10	1	10	5	50

5c: Experiment 2: Criteria Reached Day 1

		Criteria 0		Criteria 1		Criteria 2		Criteria 3		Criteria 4	
		n	%	n	%	n	%	n	%	n	%
Type of	R-/R+	13	39.39	2	6.06	0	0	1	3.03	1	3.03
Training	R-	12	36.36	2	6.06	1	3.03	0	0	1	3.03
Total:		25	75.76	4	12.12	1	3.03	1	3.03	2	6.06

5d: Experiment 2: Criteria Reached Day 3

		Criteria 0		Criteria 1		Criteria 2		Criteria 3		Criteria 4	
		n	%	n	%	n	%	n	%	n	%
Type of	R-/R+	7	22.58	0	0	2	6.45	0	0	7	22.58
Training	R-	4	12.90	3	9.68	0	0	0	0	8	25.81
Total:		11	35.48	3	9.68	2	6.45	0	0	15	48.39

Table 5. *a.* Experiment 1: Criteria reached Day 1. *b.* Experiment 1: Criteria reached Day 2. *C.* Experiment 2: Criteria Reached Day 1 *d.* Experiment 2: Criteria reached Day

a.

		Negative Change		Positive Change		No Change	
		n	%	n	%	n	%
Type of Training	Combination	12	25	9	18.75	2	4.17
	Negative	10	20.83	5	10.42	10	20.83
Total:		22	45.83	14	29.17	12	25

b

		Negative Change		Positive Change		No Change	
		n	%	n	%	n	%
Type of Training	R-/R+	15	15.31	20	20.41	15	15.31
	R-	15	15.31	8	8.16	25	25.51
Total:		30	30.61	28	28.51	40	40.82

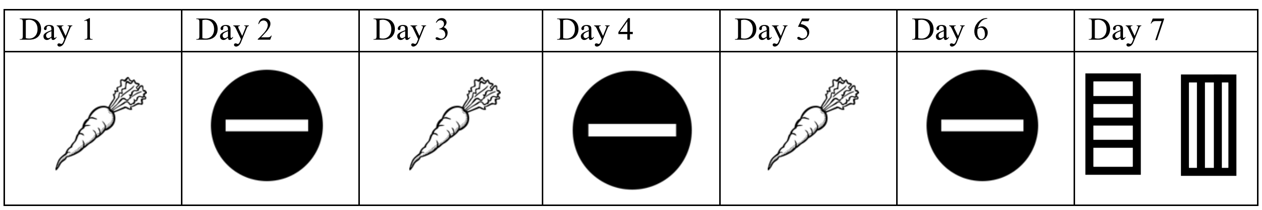
Table 6.

a. Experiment 1: Proximity difference contingency table. *b*. Experiment 2 and 3: Proximity difference contingency table.

FIGURES

Figure 1. The two-week overview schedule for horses. The carrot indicates days horses were trained with R-/R+; the negative sign indicates days horses were trained with R-. The two rectangles indicate when horses performed their preference test. *a.* Schedule for all horses in Experiment 1 and Commodore, Cruz, Del Fino, Kate Spade, Mai Tai, Riley, River, Squirt, and Kwil in Experiment 2. *b.* Schedule for all other horses in Experiment 2.

a.



b.

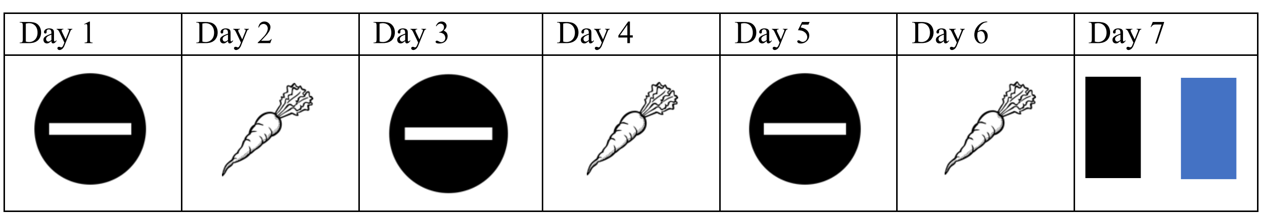


Figure 2. Schematic and dimensions of the placement of the round pen and liverpools for the overall testing area for all horses in Experiment 1 and Commodore, Cruz, Del Fino, Kate Spade, Mai Tai, Riley, River, Squirt, and Kwil in Experiment 2.

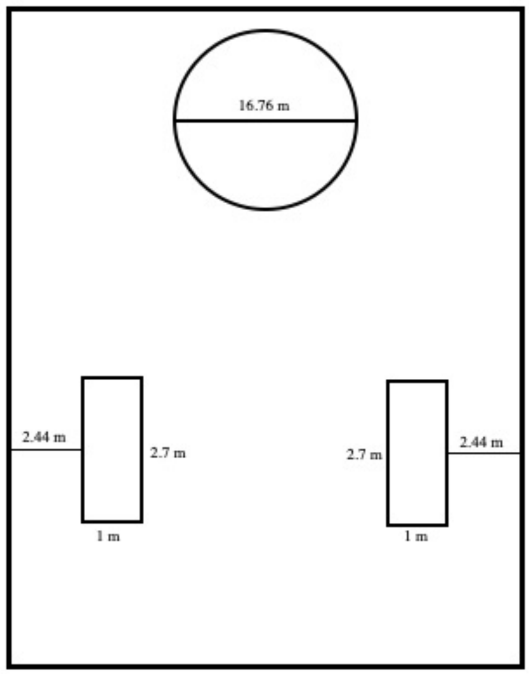


Figure 3. Dimensions of the start line from the liverpool for Experiments 1 and 2.

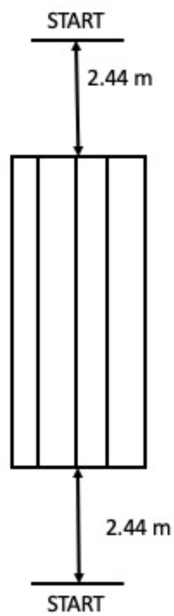
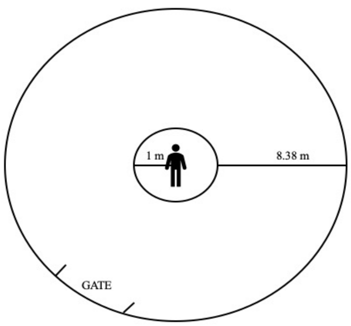
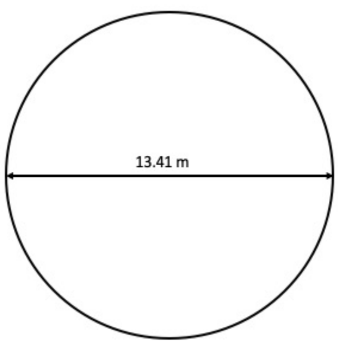


Figure 4. *a.* Schematic and dimensions of the placement of the motionless human and the round pen used for the motionless human test for Experiment 1 and some horses in Experiment 2. *b.* Schematic and dimensions of the round pen used for motionless test and preference test for all other horses in Experiment 2. *c.* Schematic and dimensions for the placement of the liverpools and horse for the preference test in Experiments 1 and 2. The horseshoe indicates the placement of the horse during the preference test.

a.



b.



c

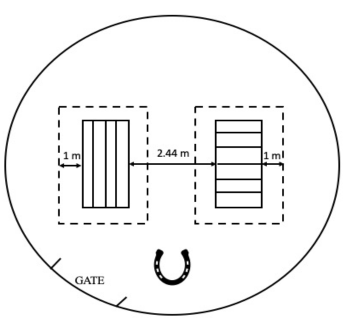
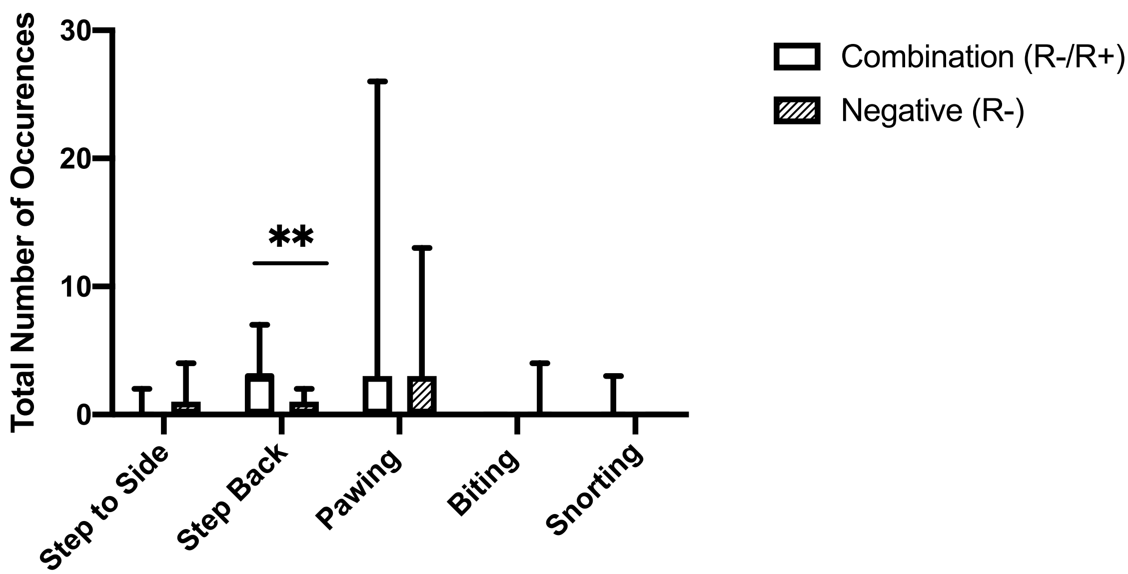


Figure 5. *a.* Experiment 1 median and IQR for occurrences of behaviors during training in each condition. *b.* Experiment 1 median and IQR for durations of behaviors during training in each condition.

a.



b.

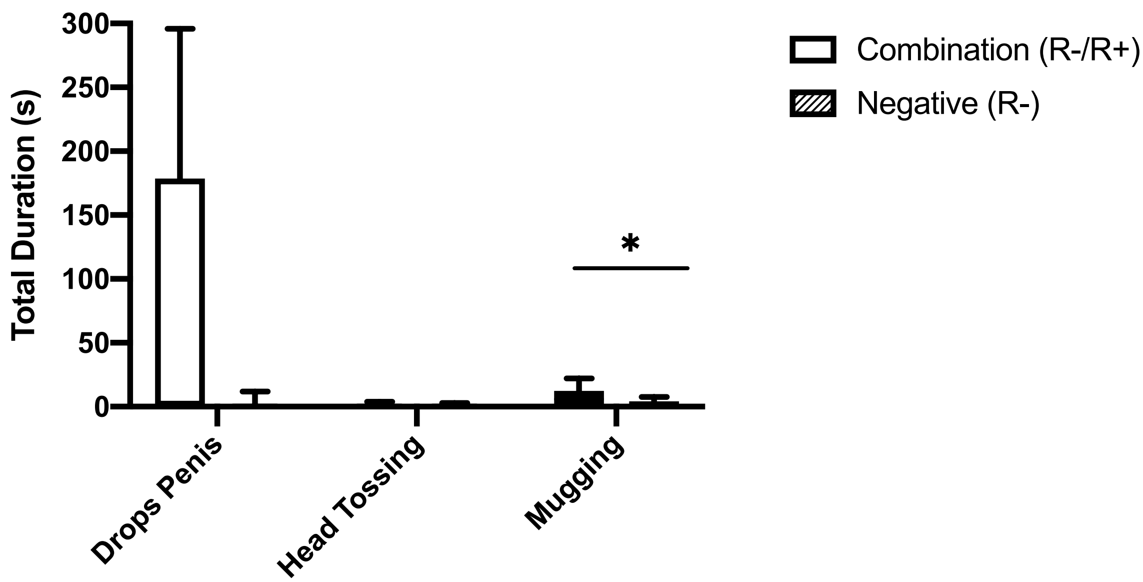
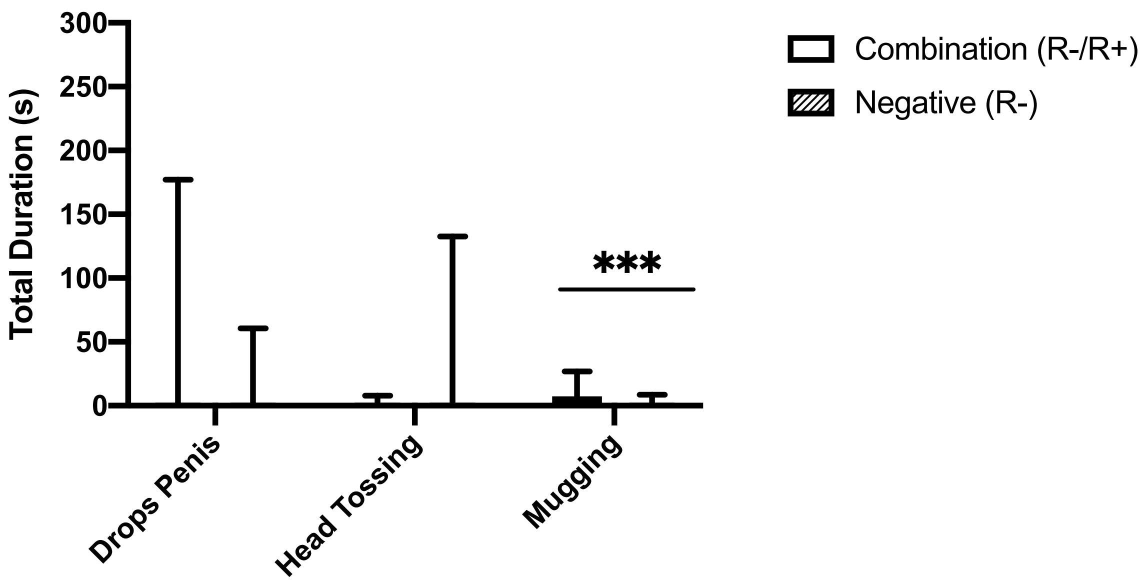


Figure 6. *a.* Experiment 2 median and IQR for occurrences of behaviors for days 2 and 3 in each condition. *b*. Experiment 2 median and IQR for durations of behaviors of behaviors for days 2 and 3 in each condition.

a.



b.

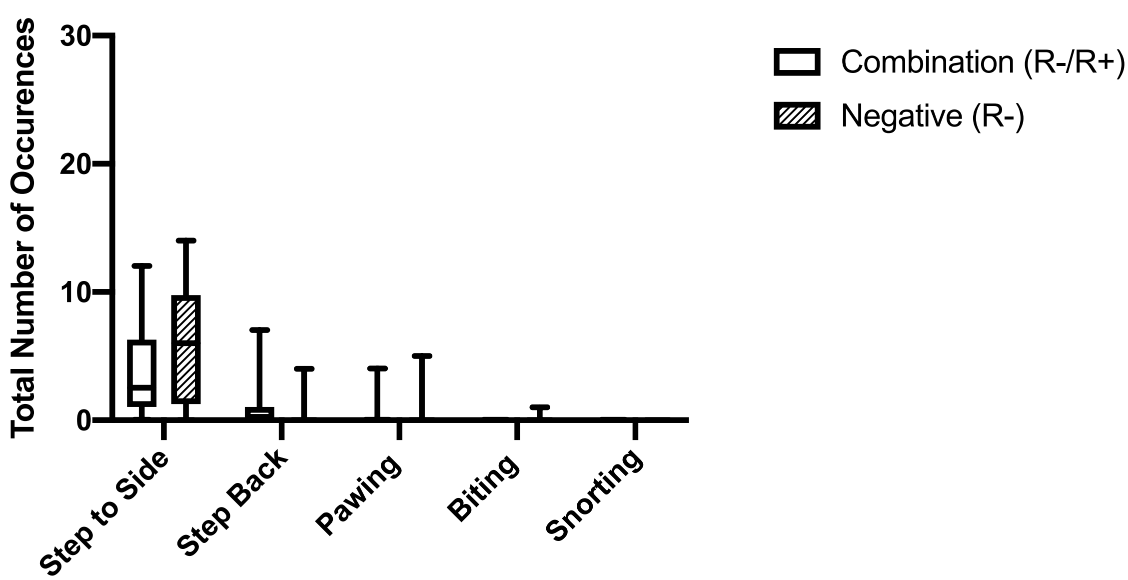
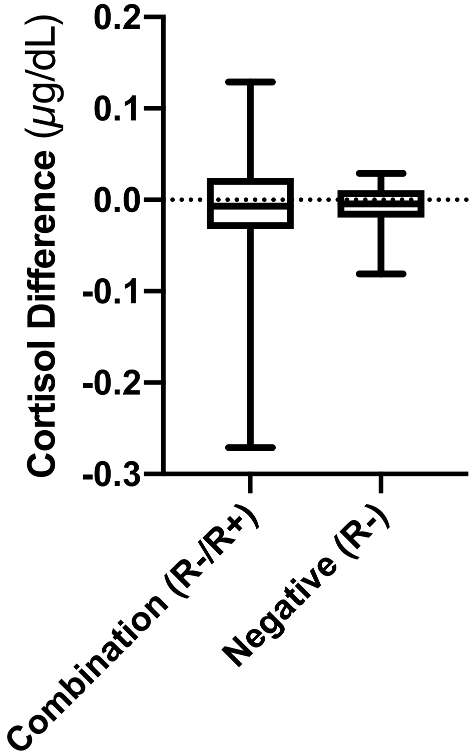
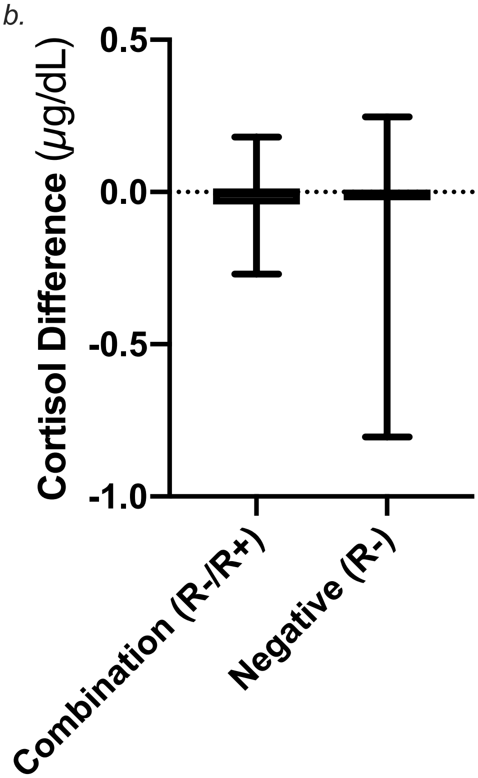


Figure 7. *a.* Experiment 1 median and IQR for cortisol difference in each condition. *b.* Experiment 2 median and IQR for cortisol difference in each condition.

a.



b.



LITERATURE CITED

Azrin, N. H., Hutchinson, R. R., & Hake, D. F. (1963). Pain-induced fighting in the squirrel monkey1. *Journal of the Experimental Analysis of Behavior*, *6*(4), 620–620.

Azrin, N. H., Hutchinson, R. R., & Sallery, R. D. (1964). Pain-aggression toward inanimate objects1. *Journal of the Experimental Analysis of Behavior*, *7*(3), 223–228.

Baragli, P., Padalino, B., & Telatin, A. (2015). The role of associative and non-associative learning in the training of horses and implications for the welfare (a review). *Annali Dell'Istituto Superiore Di Sanita*, *51*(1), 40–51.

Beerda, B., Schilder, M. B., Janssen, N. S., & Mol, J. A. (1996). The use of saliva cortisol, urinary cortisol, and catecholamine measurements for a noninvasive assessment of stress responses in dogs. *Hormones and Behavior*, *30*(3), 272–279.

Beerda, Bonne, Schilder, M. B. H., van Hooff, J. A. R. A. M., de Vries, H. W., & Mol, J. A. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, *58*(3–4), 365–381.

Bergamasco, L., Osella, M. C., Savarino, P., Larosa, G., Ozella, L., Manassero, M., Badino, P., Odore, R., Barbero, R., & Re, G. (2010). Heart rate variability and saliva cortisol assessment in shelter dog: Human–animal interaction effects. *Applied Animal Behaviour Science*, *125*(1–2), 56–68.

Cook, N. J. (2012). Review: Minimally invasive sampling media and the measurement of corticosteroids as biomarkers of stress in animals. *Canadian Journal of Animal Science*, *92*(3), 227–259.

Cooper, J. J., Cracknell, N., Hardiman, J., Wright, H., & Mills, D. (2014). The welfare consequences and efficacy of training pet dogs with remote electronic training collars in comparison to reward based training. *PloS One*, *9*(9), e102722.

d'Ingeo, S., Quaranta, A., Siniscalchi, M., Stomp, M., Coste, C., Bagnard, C., Hausberger, M., & Cousillas, H. (2019). Horses associate individual human voices with the valence of past interactions: a behavioural and electrophysiological study. *Scientific Reports*, *9*(1), 11568.

Dai, F., Dalla Costa, A., Bonfanti, L., Caucci, C., Di Martino, G., Lucarelli, R., Padalino, B., & Minero, M. (2019). Positive reinforcement-based training for self-loading of meat horses reduces loading time and stress-related behavior. *Frontiers in Veterinary Science*, *6*, 350.

Deldalle, S., & Gaunet, F. (2014). Effects of 2 training methods on stress-related behaviors of the dog (Canis familiaris) and on the dog–owner relationship. *Journal of Veterinary Behavior: Clinical Applications and Research: Official Journal of: Australian Veterinary Behaviour Interest Group, International Working Dog Breeding Association*, *9*(2), 58–65.

Desportes, G., Buholzer, L., Anderson-Hansen, K., Blanchet, M.-A., Acquarone, M., Shephard, G., Brando, S., Vossen, A., & Siebert, U. (2007). Decrease stress; Train your animals: The effect of handling methods on cortisol levels in harbour porpoises (Phocoena phocoena) under human care. *Aquatic Mammals*, *33*(3), 286–292.

Elliot, O., & King, J. A. (1960). Effect of early food deprivation upon later consummatory behavior in puppies. *Psychological Reports*, *6*(3), 391–400.

Feh, C., & de Mazières, J. (1993). Grooming at a preferred site reduces heart rate in horses. *Animal Behaviour*, *46*(6), 1191–1194.

Ferguson, D. L., & Rosales-Ruiz, J. (2001). Loading the problem loader: the effects of target training and shaping on trailer-loading behavior of horses. *Journal of Applied Behavior Analysis*, *34*(4), 409–423.

Feuerbacher, E. N., & Wynne, C. D. L. (2012). Relative efficacy of human social interaction and food as reinforcers for domestic dogs and hand-reared wolves. *Journal of the Experimental Analysis of Behavior*, *98*(1), 105–129.

Feuerbacher, E. N., & Wynne, C. D. L. (2014). Most domestic dogs (Canis lupus familiaris) prefer food to petting: population, context, and schedule effects in concurrent choice. *Journal of the Experimental Analysis of Behavior*, *101*(3), 385–405.

Feuerbacher, E. N., & Wynne, C. D. L. (2015). Shut up and pet me! Domestic dogs (Canis lupus familiaris) prefer petting to vocal praise in concurrent and single-alternative choice procedures. *Behavioural Processes*, *110*, 47–59.

Glenk, L. M., Kothgassner, O. D., Stetina, B. U., Palme, R., Kepplinger, B., & Baran, H. (2013). Therapy dogs' salivary cortisol levels vary during animal-assisted interventions. *Animal Welfare (South Mimms, England)*, *22*(3), 369–378.

Harewood, E. J., & McGowan, C. M. (2005). Behavioral and physiological responses to stabling in naive horses. *Journal of Equine Veterinary Science*, *25*(4), 164–170.

Heleski, C., Bauson, L., & Bello, N. (2008). Evaluating the addition of positive reinforcement for learning a frightening task: a pilot study with horses. *Journal of Applied Animal Welfare Science: JAAWS*, *11*(3), 213–222.

Hockenhull, J., & Creighton, E. (2010). Unwanted oral investigative behaviour in horses: A note on the relationship between mugging behaviour, hand-feeding titbits and clicker training. *Applied Animal Behaviour Science*, *127*(3–4), 104–107.

Kieson, E., Felix, C., Webb, S., & Abramson, C. I. (2020). The effects of a choice test between food rewards and human interaction in a herd of domestic horses of varying breeds and experiences. *Applied Animal Behaviour Science*, *231*(105075), 105075.

McGreevy, P. D., & McLean, A. N. (2007). Roles of learning theory and ethology in equitation. *Journal of Veterinary Behavior: Clinical Applications and Research: Official Journal of: Australian Veterinary Behaviour Interest Group, International Working Dog Breeding Association*, *2*(4), 108–118.

McGreevy, P. D., & McLean, A. N. (2009). Punishment in horse-training and the concept of ethical equitation. *Journal of Veterinary Behavior: Clinical Applications and Research: Official Journal of: Australian Veterinary Behaviour Interest Group, International Working Dog Breeding Association*, *4*(5), 193–197.

McLean, A. N., & Christensen, J. W. (2017). The application of learning theory in horse training. *Applied Animal Behaviour Science*, *190*, 18–27.

Merali, Z., McIntosh, J., Kent, P., Michaud, D., & Anisman, H. (1998). Aversive and appetitive events evoke the release of corticotropin-releasing hormone and bombesin-like peptides at the central nucleus of the amygdala. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(12), 4758–4766.

Olczak, K., Winther Christensen, J., & Klocek, C. (2018). Food motivation in horses appears stable across different test situations. *Applied Animal Behaviour Science*, *204*, 60–65.

Padalino, B., Maggiolino, A., Boccaccio, M., & Tateo, A. (2012). Effects of different positions during transport on physiological and behavioral changes of horses. *Journal of Veterinary Behavior: Clinical Applications and Research: Official Journal of: Australian Veterinary*

Behaviour Interest Group, International Working Dog Breeding Association, *7*(3), 135–141.

Rescorla, R. A. (1967). Pavlovian conditioning and its proper control procedures. *Psychological Review*, *74*(1), 71–80.

Sankey, C., Henry, S., Clouard, C., Richard-Yris, M.-A., & Hausberger, M. (2011). Asymmetry of behavioral responses to a human approach in young naive vs. trained horses. *Physiology & Behavior*, *104*(3), 464–468.

Sankey, C., Henry, S., Górecka-Bruzda, A., Richard-Yris, M.-A., & Hausberger, M. (2010). The way to a man's heart is through his stomach: what about horses? *PloS One*, *5*(11), e15446.

Sankey, C., Richard-Yris, M.-A., Henry, S., Fureix, C., Nassur, F., & Hausberger, M. (2010). Reinforcement as a mediator of the perception of humans by horses (Equus caballus). *Animal Cognition*, *13*(5), 753–764.

Sankey, C., Richard-Yris, M.-A., Leroy, H., Henry, S., & Hausberger, M. (2010). Positive interactions lead to lasting positive memories in horses, Equus caballus. *Animal Behaviour*, *79*(4), 869–875.

Schilder, M. B. H., & van der Borg, J. A. M. (2004). Training dogs with help of the shock collar: short and long term behavioural effects. *Applied Animal Behaviour Science*, *85*(3–4), 319–334.

Starling, M., McLean, A., & McGreevy, P. (2016). The contribution of Equitation Science to minimising horse-related risks to humans. *Animals: An Open Access Journal from MDPI*, *6*(3), 15.

Thorbergson, Z. W., Nielsen, S. G., Beaulieu, R. J., & Doyle, R. E. (2016). Physiological and behavioral responses of horses to wither scratching and patting the neck when under saddle. *Journal of Applied Animal Welfare Science: JAAWS*, *19*(3), 245–259.

Valenchon, M., Lévy, F., Moussu, C., & Lansade, L. (2017). Stress affects instrumental learning based on positive or negative reinforcement in interaction with personality in domestic horses. *PloS One*, *12*(5), e0170783.

Vicars, S. M., Miguel, C. F., & Sobie, J. L. (2014). Assessing preference and reinforcer effectiveness in dogs. *Behavioural Processes*, *103*, 75–83.

Wathan, J., & McComb, K. (2014). The eyes and ears are visual indicators of attention in domestic horses. *Current Biology: CB*, *24*(15), R677-9.

Williams, J. L., Friend, T. H., Nevill, C. H., & Archer, G. (2004). The efficacy of a secondary reinforcer (clicker) during acquisition and extinction of an operant task in horses. *Applied Animal Behaviour Science*, *88*(3–4), 331–341.

York, A., Matusiewicz, J., & Padalino, B. (2017). How to minimise the incidence of transport-related problem behaviours in horses: a review. *Journal of Equine Science*, *28*(3), 67–75.

Carter, D. A., & Lightman, S. L. (1987). Oxytocin responses to stress in lactating and hyperprolactinaemic rats. *Neuroendocrinology*, *46*(6), 532–537.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2014). *Applied behavior analysis: Pearson new international edition* (2nd ed.). Pearson Education.

Lansade, L., Nowak, R., Lainé, A.-L., Leterrier, C., Bonneau, C., Parias, C., & Bertin, A. (2018). Facial expression and oxytocin as possible markers of positive emotions in horses. *Scientific Reports*, *8*(1), 14680.

McCall, C. A., & Burgin, S. E. (2002). Equine utilization of secondary reinforcement during response extinction and acquisition. *Applied Animal Behaviour Science*, *78*(2–4), 253–262.

Sankey, C., Henry, S., Górecka-Bruzda, A., Richard-Yris, M.-A., & Hausberger, M. (2010). The way to a man's heart is through his stomach: what about horses? *PloS One*, *5*(11), e15446.

Sankey, C., Richard-Yris, M.-A., Henry, S., Fureix, C., Nassur, F., & Hausberger, M. (2010). Reinforcement as a mediator of the perception of humans by horses (Equus caballus). *Animal Cognition*, *13*(5), 753–764.

Sankey, C., Richard-Yris, M.-A., Leroy, H., Henry, S., & Hausberger, M. (2010). Positive interactions lead to lasting positive memories in horses, Equus caballus. *Animal Behaviour*, *79*(4), 869–875.

Waran, N., McGreevy, P., & Casey, R. A. (2007). Training methods and horse welfare. In *Animal Welfare* (pp. 151–180). Springer Netherlands.